Unravelling the Solution Structure of Secretory Immunoglobulin A

Thesis Presented for the Degree of

Doctor of Philosophy

by

Alexandra Bonner

Institute of Structural and Molecular Biology,

University College London.

March 2008

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

I, Alexandra Bonner, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Immunoglobulin A (IgA) is the most abundant and heterogenous human antibody class, being present in two major subclasses, IgA1 and IgA2, and different oligomeric states. Secretory component (SC) binds to dimeric IgA1 (dIgA) to form secretory IgA (SIgA). In this thesis, X-ray and neutron scattering, analytical ultracentrifugation and constrained modelling were used to determine their structures in solution. Free SC consists of five Ig-variable domains, with seven glycosylation sites. SC was found to adopt a compact, J-shaped structure in solution, with the glycosylation sites probably located on one face of SC. In this model, one side of SC and SC D1, containing known IgA-binding motifs, is free to interact with dIgA. Dimeric IgA1 (dIgA1) is formed by two IgA1 monomers covalently bound by the joining (J) chain. The modelling of the solution structure of dIgA1 shows that the arrangement of the two monomers are near-planar in an extended arrangement. The two Fc regions form a slightly bent end-to-end contact. All four Fab antigen-binding sites are independent of one another. The model shows that the two Fc regions are accessible to the Fc alpha receptor (FcαRI) and the polymeric immunoglobulin receptor (pIgR). This dIgA1 structure suggests a mechanism for dIgA1 aggregation in IgA nephropathy through oligomerisation caused by removal or movement of the Fab regions by proteolysis or changes in the glycosylation of the hinge region. SIgA is formed by the transcytosis of dIgA across epithelial cells, where after binding pIgR (membrane-bound SC) it is secreted into the mucosa. At the mucosal surfaces SIgA acts as the first line of defence against pathogens. The solution structure of SIgA1 suggests that the arrangement of the IgA1 monomers are not altered when SC is binds to dIgA1, therefore retaining their near-planar structure. The five domains of SC are in an extended conformation along the convex Fc-Fc edge of the dimer. The SIgA1 solution structure shows how SC and dIgA1 confer mutual protection to one another from proteolysis. The modelling of the solution structure of SIgA2 shows that, although the IgA2 monomers are in a similar arrangment to that in SIgA1, SIgA2 has a non-planar structure in solution. SC is located in a similar extended arranagment on the Fc-Fc convex edge to that in SIgA1, yet the Fab regions are positioned out of the Fc-Fc plane. These structural differences between the SIgA isotypes may increase their efficiency in mucosal immunity but confer varied susceptibilities to mucosal proteases.

Acknowledgements

I would firstly like to thank my supervisor Professor Stephen J. Perkins for his support and encouragement, and many helpful discussions during the course of my work with him.

I would like to thank my collaborators, including Professor Blaise Corthésy, Dr. Clementine Perrier, Professor Michael Kerr and Dr. Adel Almogren for their expertise and for supplying the proteins used for my studies. Thanks also to the scientists providing instrumental support, particularly Dr. Stephanie Finet, Dr. Pierre Panine, Dr. Richard Heenan and Mr Jayesh Gor.

I would like to thank Dr. Patricia Furtado for all her help, advice, friendship and her Fab enthusiasm about IgA! Thanks also to Shreenal Patel and Dr. Bex Saunders for their continual support and encouragement, and above all, their friendship. Shree, thanks for the insightful discussions about our work, and much-needed tea-breaks! I could not have got through my work without you. Bex, thanks for all your help and advice from the outset and for being my favorite drinking buddy!

Alice Trussler, my twin, thank you for being the best-friend that anyone could wish for. You've always been there for me and I really appreciate all our long chats (with bottle of wine or two!) and your advice.

I would like my Mum and Dad, Simon and Sharon to know that I could not have done any of this without them. The daunting task of getting this far in the first place and writing my thesis has been attainable only with such strong backing, from people who do make you believe that you can do anything.

Finally, I would like to thank Dean, for his patience, unfaltering support, humour and his love.

Contents **Page**

Chapter One

Biochemistry of Secretory Immunoglobulin A 1

1.1 The adaptive immune system 2

1.1.1 The innate and adaptive immune system 2

1.1.2 Introduction to lymphocytes of the adaptive immune system 2

1.1.3 The humoral immune system 3

1.1.4 B-cell maturation and effector functions 5

1.2 Antibodies 7

1.2.1 General antibody structure 7

1.2.2 Immunoglobulin domains 10

1.2.2.1 Ig domain common core and intradomain disulphide bonds 13

1.2.2.2 Complementarity Determining Regions 13

1.2.3 Fab arm and antigen-binding site formation 14

1.2.4 Antibody classes and biological activities 16

1.2.4.1 Immunoglobulin G 18

1.2.4.2 Immunoglobulin M 18

1.2.4.3 Immunoglobulin D 20

1.2.4.4 Immunoglobulin E 21

1.3 Immunoglobulin A 21

1.3.1 Biochemistry of monomeric IgA 24

1.3.1.1 N- and O-linked carbohydrates 26

1.3.2 Overview of the interaction of IgA with other proteins 27

1.3.3 Dimeric IgA 28

1.3.3.1 Joining chain 28

1.4 Secretory IgA 32

1.4.1 Formation and secretion of SIgA 32

1.4.2 Secretory component / polymeric immunoglobulin receptor 35

1.4.2.1 Structure of SC domains and their interaction with SIgA 36

1.4.2.2 Functions of free SC 38

1.4.2.3 Human breast milk and colostrum 38

1.4.3 IgM as an accessory secretory immunoglobulin 39

1.5 Mucosal immunity and the SIgA response **39**
1.5.1 Mucosal Associated Lymphreticular Tissue (MALT) 40
1.5.2 The role of SIgA in mucosal defence 42
1.5.3 Mucosal microflora and SIgA in mucosal tolerance 43
1.6 IgA receptors **44**
1.6.1 IgA Fcα receptor 44
1.6.2 IgA Fcα/μ receptor 47
1.6.3 Secretory component receptor 48
1.6.4 Transferrin receptor 48
1.6.5 Asialoglycoprotein receptor 48
1.7 Immunoglobulin A in disease **49**
1.7.1 Selective IgA deficiency 49
1.7.2 IgA nephropathy (IgAN) 49
1.7.3 Other diseases associated with IgA 51
1.8 Therapeutic uses of IgA **52**

Chapter Two
Methods to Determine Antibody Structure **54**
2.1 Introduction antibody structure determination methods **55**
2.1.1 Electron Microscopy 56
2.1.2 Nuclear Magnetic Resonance 57
2.1.3 X-ray crystallography 57
2.2 Solution scattering **58**
2.2.1 Complementarity of solution scattering with other methods 59
2.2.2 Theory of solution scattering 59
2.2.2.1 The Debye equation 63
2.2.2.2 The two-phase model of solution scattering 65
2.2.2.3 Contrast variation 66
2.2.2.4 The hydration shell 66
2.2.3 Solution scattering methods 67
2.2.3.1 Sample preparation 67
2.2.3.2 X-ray Data collection on instrument ID02 at the ESRF 68

2.2.3.3 Neutron data collection on instrument LOQ at ISIS 71
2.2.4 Analyses of reduced scattering curves $I(Q)$ 73
2.2.4.1 Guinier analysis 73
2.2.4.2 The cross-sectional radius of gyration 74
2.2.4.3 R_G and R_{XS} determination of macromolecular dimensions 76
2.2.4.4 The real space distance distribution function 76
2.3 Analytical ultracentrifugation **77**
2.3.1 Absorption optics 79
2.3.2 Interference optics 79
2.3.3 Sedimentation velocity 79
2.3.4 Sedimentation equilibrium 83
2.4 Combining solution scattering and AUC **83**

Chapter Three
Constrained Biomolecular Modelling **85**
3.1 Introduction to Protein structure **86**
3.1.1 Amino acids 86
3.1.2 Secondary structure 88
3.1.2.1 Peptide bonds and Ramachandran Plots 88
3.1.2.2 Alpha-helices and Beta-sheets 88
3.1.2.3 Domain structure 90
3.2 Biomolecular modelling **90**
3.2.1 Homology modelling 91
3.2.1.1 Limitations of homology modelling and assessing the structure 92
3.3 Constrained scattering modelling **92**
3.3.1 Analysis of glycoprotein composition 93
3.3.2 Creating atomic models 93
3.3.3 Debye sphere modelling 94
3.3.4 Debye scattering curve calculation 96
3.3.5 Evaluation of models 97

Chapter Four

Solution Structure of Secretory Component: Implications for Biological Function 99

4.1 Introduction 100

4.2 Results **103**

4.2.1 Characterisation of SC and its fragments 103

4.2.2 X-ray and neutron Guinier analyses of SC and its fragments 103

4.2.3 X-ray and neutron distance distribution analyses of SC and its fragments 110

4.2.4 Analytical ultracentrifugation analyses of SC and its fragments 113

4.2.5 Constrained modelling of SC and its fragments 119

4.3 Discussion **131**

4.4 Materials and methods **136**

4.4.1 Preparation of recombinant SC and its fragments 136

4.4.2 X-ray and neutron scattering data collection 137

4.4.3 Sedimentation equilibrium and sedimentation velocity data acquisition 138

4.4.4 Conformational searches for the best-fit solution model of SC 139

4.4.5 Debye scattering and sedimentation coefficient modelling of SC 140

4.4.6 SC Protein Data Bank accession code 141

Chapter Five

Implications of the Near-Planar Solution Structure of Human Myeloma Dimeric IgA1 for Mucosal Immunity and IgA Nephropathy **142**

5.1 Introduction 143

5.2 Results **144**

5.2.1 Purification of human myeloma dIgA1 144

5.2.2 X-ray and neutron scattering analyses of dIgA1 144

5.2.3 Analytical ultracentrifugation of dIgA1 153

5.2.4 dIgA1 constrained scattering modelling 158

5.3 Discussion **174**

5.4 Materials and methods **183**

5.4.1 Preparation and composition of dIgA1 183

5.4.2 X-ray and neutron scattering data collection and analysis 183

5.4.3 Analytical ultracentrifugation data collection and analysis for dIgA1 184
5.4.4 Constrained modelling of dIgA1 185
5.4.5 dIgA1 and pentameric IgM Protein Data Bank accession codes 186

Chapter Six

Location of Secretory Component on the Fc Edge of Dimeric IgA1 Reveals Insight into the Role of Secretory IgA1 in Mucosal Immunity **187**
6.1 Introduction **188**
6.2 Results **190**
6.2.1 Purification of colostrum SIgA1 190
6.2.2 X-ray and neutron scattering of SIgA1 191
6.2.3 SIgA1 analytical ultracentrifugation analyses 195
6.2.4 Constrained modelling of SIgA1 198
6.3 Discussion **208**
6.4 Materials and Methods **216**
6.4.1 Preparation and composition of SIgA1 216
6.4.2 X-ray and neutron scattering and ultracentrifugation data 216
6.4.3 Constrained modelling of SIgA1 218
6.4.4 SIgA1 Protein Data Bank accession code 220

Chapter Seven

Solution Structure of Human SIgA2: Implications for Mucosal Immunity of the Non-Planar IgA2 and Near-Planar IgA1 solution structures in human Secretory IgA **221**
7.1 Introduction **222**
7.2 Results **225**
7.2.1 Purification of human colostrum SIgA2 225
7.2.2 X-ray and neutron scattering of SIgA2 226
7.2.3 Analytical ultracentrifugation of SIgA2 230
7.2.4 Constrained homology modelling of SIgA2 233
7.3 Discussion **245**
7.4 Materials and Methods 250
7.4.1 Preparation and composition of SIgA2 250

7.4.2 X-ray and neutron scattering and ultracentrifugation data 251
7.4.3 Constrained modelling of SIgA2 252
7.4.4 SIgA2 Protein Data Bank accession code 254

Chapter Eight
Conclusions **255**
8.1 Conclusions **256**
8.1.1 History of structural models for secretory IgA 256
8.1.2 Solution structures of SC, dIgA and the SIgA isoforms 259

References **265**
Publications **280**
Presentations and abstracts **281**

Contents- Figures **Page**

Chapter One

Figure 1.1 Overview of the humoral immune system 4

Figure 1.2 B cell maturation 6

Figure 1.3 Distribution of effector sites of the different antibody classes throughout the body 8

Figure 1.4 Generic antibody structure 9

Figure 1.5 Two-dimensional topology diagrams for each of the four Ig folds 11

Figure 1.6 Location of the complementarity determining regions (CDRs) in variable immunoglobulin domains of antibody light chains 15

Figure 1.7 Electron micrograph and solution structure of pentameric IgM 19

Figure 1.8 Solution structure of IgA1 and IgA2m(1) 23

Figure 1.9Electron micrographs of dimeric IgA and secretory IgA 29

Figure 1.10 Schematic structure of dimeric/secretory IgA 30

Figure 1.11 Transportation of dimeric IgA across the epithelial lining and formation of secretory IgA 33

Figure 1.12 Structure of the polymeric immunoglobulin receptor 34

Figure 1.13 Crystal structure of secretory component domain one 37

Figure 1.14 Antigen processing and mucosal immune response induction 41

Figure 1.15 Interaction between the Fc regions of IgA and FcαRI 45

Chapter Two

Figure 2.1 Schematic representation of a scattering experiment 60

Figure 2.2 General features of an X-ray solution scattering curve 64

Figure 2.3 X-ray solution scattering at the ESRF 69

Figure 2.4 Neutron solution scattering at ISIS 72

Figure 2.5 Molecular mass against I(0)/c for proteins measured on LOQ 75

Figure 2.6 Schematic diagram of AUC optic systems 78

Figure 2.7 Schematic and bitmap image showing SV and SE boundaries 81

Figure 2.8 SEDFIT analysis output 82

Chapter Three

Figure 3.1 The 20 amino acid side chains 87

Figure 3.2 Peptide bond angles 89

Figure 3.3 Overview of the constrained modelling procedure 95

Chapter Four

Figure 4.1 The domain structure of SC and its sequence alignment 101

Figure 4.2 SDS-PAGE and Western blot analyses of SC and its fragments 104

Figure 4.3 Guinier R_G and R_{XS} analyses for SC 106

Figure 4.4 Guinier R_G and R_{XS} analyses for D1-D3 and D4-D5 107

Figure 4.5 X-ray and neutron distance distribution function *P(r)* for SC 111

Figure 4.6 X-ray and neutron distance distribution function *P(r)* for D1-D3 and D4-D5 112

Figure 4.7 Sedimentation equilibrium fits for SC 114

Figure 4.8 SEDFIT sedimentation velocity fits for SC 116

Figure 4.9 SEDFIT sedimentation velocity fits for D1-D3 and D4-D5 117

Figure 4.10 DCDT+ sedimentation velocity fits for SC and D4-D5 118

Figure 4.11 Constrained scattering analyses of the X-ray and neutron data for D1-D3 and D4-D5 122

Figure 4.12 Best-fit and poor-fit D1-D3 and D4-D5 models for the X-ray and neutron data 123

Figure 4.13 Constrained scattering analyses of the SC X-ray and neutron data 126

Figure 4.14 Best-fit and poor-fit SC models for the X-ray and neutron data 128

Figure 4.15 Comparison of the solution scattering models for monomeric IgA1 and IgA2 with the best-fit models of SC and its D1-D3 and D4-D5 fragments 129

Figure 4.16 Ten best-fit models of SC 130

Figure 4.17 Detailed views of the best-fit model of SC 133

Chapter Five

Figure 5.1 The domain structure of dIgA1 145

Figure 5.2 The sequence of an IgA1 monomer and J chain 146

Figure 5.3 Purification of human myeloma dIgA1 147

Figure 5.4 X-ray and neutron Guinier analyses for dIgA1 149

Figure 5.5 X-ray and neutron distance distribution functions *P(r)* for dIgA1 152

Figure 5.6 Sedimentation equilibrium analyses for dIgA1 154

Figure 5.7 dIgA1 sedimentation velocity *c(s)* distribution analyses using SEDFIT 156

Figure 5.8 Time-derivation sedimentation velocity analyses of dIgA1 using DCDT+ 157

Figure 5.9 The starting model for the dIgA1 fit-cycle one searches and the orientations of the best-fit models for dIgA1 160

Figure 5.10 Outcome of the dIgA1 modelling from fit-cycle one 162

Figure 5.11 X-ray and neutron scattering curve fits for the best-fit model for dIgA1 and three poor-fit models from fit-cycle one 163
Figure 5.12 Best-fit models for dIgA1 from fit-cycle one 164
Figure 5.13 Effect of adding J chain to dIgA1 models from Search 1 of modelling fit-cycle one 166
Figure 5.14 Effect of adding human IgA1 Fc crystal structure to the best-fit dIgA1 structure from fit-cycle one 168
Figure 5.15 Outcome of the dIgA1 modelling from fit-cycle two 169
Figure 5.16 Scattering curves for the best-fit dIgA1 model from fit-cycle two 170
Figure 5.17 Superimposition of the 32 best-fit dIgA1 structure from fit cycle two 171
Figure 5.18 Ten best-fit dIgA1 structures from fit-cycle two 173
Figure 5.19 molecular architecture of the dIgA1 solution structure 175
Figure 5.20 Interaction of dIgA1 with two FcαRI molecules 179

Chapter Six

Figure 6.1 Domain structure of SIgA1 189
Figure 6.2 SIgA1 Guinier and distance distribution functions *P(r)* analyses 192
Figure 6.3 Sedimentation equilibrium analyses for SIgA1 196
Figure 6.4 Sedimentation velocity analyses for SIgA1 using SEDFIT *c(s)* distribution and DCDT+ time-derivative fits 197
Figure 6.5 Comparison of the SIgA1 and dIgA1 experimental *I(Q)* scattering curves 199
Figure 6.6 Constrained modelling Search 1 analysis of SIgA1 200
Figure 6.7 Analyses of all ten searches from the first-round of the SIgA1 constrained modelling 201
Figure 6.8 X-ray and neutron fits for the best-fit SIgA1 model from Search 1 and two poor-fit SIgA1 models 204
Figure 6.9 Analysis of the SIgA1 fab reorientation second-round modelling search 206
Figure 6.10 Superimposition of the best 50 SIgA1 models from the Fab-refinement search 207
Figure 6.11 Superimposition of the 34 best-fit SIgA1 models from Search 1 209
Figure 6.12The ten best-fit SIgA1 models from Search 1 210
Figure 6.13 Stereo view of the best-fit SIgA1 solution structure and its interaction with an FcαRI molecule 213

Chapter Seven

Figure 7.1 Domain structure of SIgA2 223

Figure 7.2 Sequence alignment of IgA1 and the IgA2 allotypes 224

Figure 7.3 Guinier plots and distance distribution functions *P(r)* analyses for SIgA2 227

Figure 7.4 Sedimentation equilibrium analyses for SIgA2 231

Figure 7.5 Sedimentation velocity analyses for SIgA2 using SEDFIT *c(s)* distribution and DCDT+ time-derivative fits 232

Figure 7.6 Constrained modelling Search 1 analysis of SIgA2 235

Figure 7.7 Analyses of all ten searches from the first-cycle of the SIgA2 modelling 236

Figure 7.8 X-ray and neutron fits ofr the best-fit SIgA2 model and three representative poor-fit SIgA2 models 239

Figure 7.9 Analysis of the SIgA2 Fab reorientation search from the second-cycle of modelling fits 240

Figure 7.10 Superimposition of the 31 best-fit models for SIgA2 from the Fab reorientation search from the second-cycle of modelling fits 241

Figure 7.11 Superimposition of the 50 best-fit SIgA2 models from Search 1 243

Figure 7.12 The ten best-fit SIgA2 models from Search 1 244

Figure 7.13 Stereo view of the solution structure of SIgA2 246

Chapter Eight

Figure 8.1 History of structural models for secretory IgA 257

Figure 8.2 Solution structure of SC and dIgA1 260

Figure 8.3 Solution structure of SIgA1 262

Figure 8.4 Solution structure of SIgA2 264

Content-Tables **Page**

Chapter One

Table 1.1 Antibody bioactivities 17

Table 1.2 Concentrations of IgA in serum and human secretions and comparison to that of IgG 22

Chapter Two

Table 2.1 Scattering lengths f (X-rays) and b (neutrons) of biologically important atoms 62

Table 2.2 Scattering densities of solvents and biological macromolecules 62

Chapter Four

Table 4.1 Experimental X-ray and neutron scattering and AUC analyses for SC, D1-D3 and D4-D5 108

Table 4.2 Summary of the X-ray and neutron modelling fits for the SC, D1-D3 and D4-D5 solution structures 121

Chapter Five

Table 5.1 Comparison of the Guinier and P(r) analyses for human dIgA1 and three forms of monomeric IgA 150

Table 5.2 Summary of the X-ray and neutron modeling fits for the dIgA1 solution structure 159

Chapter Six

Table 6.1 Comparison of the X-ray and neutron data analyses for SIgA1, dIgA1 and monomeric IgA1 193

Table 6.2 Summary of the X-ray and neutron modeling fits from Search 1 and the Fab reorientation search for the SIgA1 solution structure 203

Chapter Seven

Table 7.1 Comparison of X-ray and neutron data analyses for the polymeric forms of human IgA and its secretory component 228

Table 7.2 Summary of the X-ray and neutron modelling fits for the SIgA2 solution structure 238

Abbreviations

Ab	antibody
Ag	antigen
ADCC	antibody-dependent cell cytotoxicity
APC	antigen-presenting cell
ASGP-R	asialoglycoprotein receptor
AUC	analytical ultracentrifugation
BALT	bronchial-associated lymphoid tissues
BM	bending magnets
C	constant (-type Ig domain)
cbpA	choline binding protein A
cDNA	complementary DNA (deoxyribonucleic acid)
CDR	complementarity determining region
CHO	Chinese hamster ovary
CTL	cytotoxic T lymphocytes
DC	dendritic cells
DSSP	dictionary of secondary structure of proteins
dIgA	dimeric immunoglobulin A
dIgA1	dimeric immunoglobulin A1
dIgA2	dimeric immunoglobulin A2
EM	electron microscopy
ESRF	European Synchrotron Radiation Facility
GALT	gut-associated lymphoid tissue
ID	insertion device
Ig	immunoglobulin
IgA	immunoglobulin A
IgD	immunoglobulin D
IgE	immunoglobulin E
IgG	immunoglobulin G
IgM	immunoglobulin M
IL	interleukin
ITAM	immunoreceptor tyrosine-based activation motif
J chain	joining chain

L	length
M	microfold (cells)
MALT	mucosal-associated lymphoid tissue
MBL	mannan-binding lectin
MHC	membrane histocompatibility complex
M_r	molecular mass
NMR	nuclear magnetic resonance
PBS	phosphate buffered saline
PP	Peyer's patch
PDB	protein data bank
pIgR	polymeric immunoglobulin receptor
R_G	radius of gyration
r.m.s.d	root mean square deviation
RNA	ribonucleic acid
R_{XS}	cross-sectional radius of gyration
S	Svedburg
SANS	small-angle neutron scattering
SAS	small-angle scattering
SAXS	small-angle X-ray scattering
SC	secretory component
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SE	sedimentation equilibrium
SIgA	secretory immunoglobulin A
SIgA1	secretory immunoglobulin A1
SIgA2	secretory immunoglobulin A2
SV	sedimentation velocity
T_C	cytotoxic T cells
TfR	transferring receptor
T_H	helper T cells
TGF	transforming growth factor
V	variable (-type Ig domain)

Chapter One

Biochemistry of Secretory Immunoglobulin A

1.1 The adaptive immune system

1.1.1 The innate and adaptive immune system

The immune system is a versatile and complex defence system that has evolved to protect animals from a plethora of pathogens, such as bacteria, viruses, other microorganisms and also potentially dangerous host-cells. The human immune system can be divided into the innate and adaptive immune system. However, it is the ability of these two branches of the immune system to co-operate and interact with one another that makes the human immune system so highly regulated and competent in its elimination of pathogens. The co-ordination of the innate and adaptive immune system can produce a more effective response when combined.

The innate immune system provides the first line of defence to an infection by using anatomic, physiologic, phagocytic and inflammatory defensive barriers. Innate immunity is more primitive and less specific than adaptive immunity. Most components of the innate immune system are present prior to an infection and are involved in pathogen-resistance mechanisms that eliminate invading organisms in a non-specific, generic way.

The adaptive immune system is highly evolved to invoke a specific immune response involving the production of immune system cells and antibodies against a particular pathogen. Throughout its life span an individual can adapt the response to an infection caused by a particular pathogen. The first encounter to an antigenic challenge may typically take up to six days, whereas a second response to the same antigen in the future may take only two days, producing a rapid and heightened immune response. This is because unlike innate immunity, adaptive immunity has immunologic memory, antigenic specificity, diversity and the ability to recognise self from non-self.

1.1.2 Lymphocytes of the adaptive immune system

The two major populations of lymphocytes are B cells and T cells. B cells mature in the bone marrow (defined after their identification in the bursa of Fabricus, the site of haematopoiesis in birds) and T cells mature in the thymus gland. These B and T cells mediate and define the attributes of the adaptive immune system.

Mature T lymphocytes express a unique, specific membrane-bound T cell receptor that will only recognise an antigen (Ag) when bound to cell membrane proteins called membrane histocompatibility complex (MHC) molecules. These MHC glycoproteins are genetically diverse and function in this Ag recognition process, known as Ag-presentation. MHC molecules are either class I or class II, which are expressed on different cell lines; MHC class I are expressed on most nucleated cells, and MHC class II are only present on Ag-presenting cells (APCs). After interaction with an Ag-MHC complex, the T cells differentiate and proliferate into effector cells to carry out specific functions. There are two subclasses of effector T cells; helper T cells (T_H) and cytotoxic T cells (T_C) that are distinguished from one another depending on the type of membrane-bound accessory receptor expressed on their cell surface. T_H cells express CD4 glycoproteins and are generally activated by Ags in complex with MHC class II molecules. Effector T_H cells secrete specific cytokines that will subsequently activate T_C cells, B cells and macrophages. After activation by T_H cells, T_C cells recognise MHC class I bound Ags, and then differentiate into cytotoxic T lymphocyte (CTL) effector cells, which perform cell-killing activity. The CTL has an essential role in monitoring and consequently eliminating any altered self-cells or those that display MHC class I bound Ag.

APCs include B cells, macrophages and dendritic cells. APCs express class II MHC molecules and also deliver a co-stimulatory signal to activate and carefully control T_H cells.

B cells express a specific B cell receptor, which is a membrane-bound antibody (Ab). When a naive B cell first encounters a specific Ag that can interact with its membrane-bound Ab, it rapidly proliferates and the progeny differentiate into memory B cells or effector B cells called plasma cells (Section 1.1.4).

1.1.3 The humoral immune system

The adaptive immune system can be further subdivided into humoral and cell-mediated immune responses (Figure 1.1). Cell-mediated immunity is the response generated by effector T cells, and includes cytokines released by T_H cells that induce phagocytic cells and CTL cells eradicating altered self-cells.

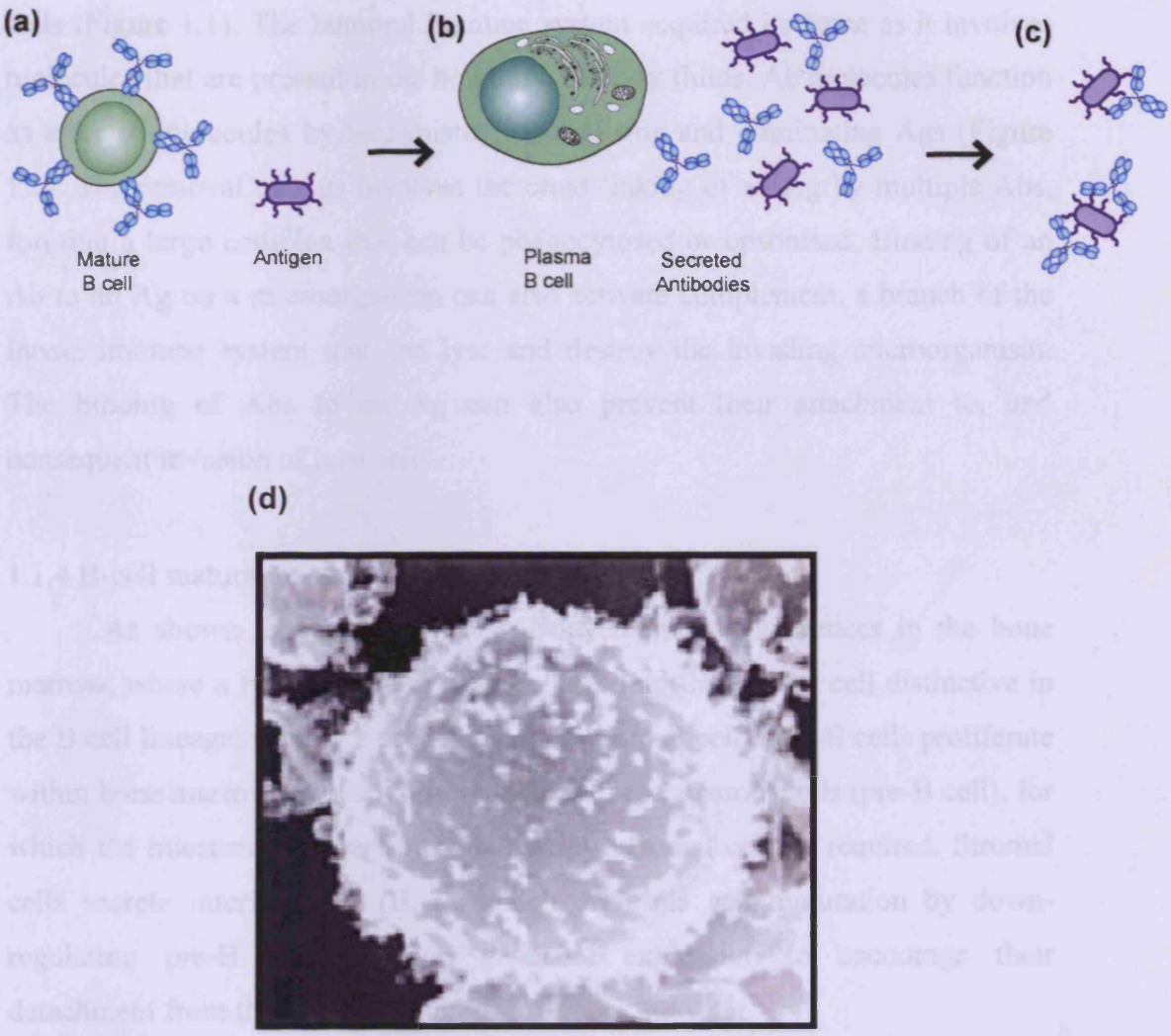

Figure 1.1 Overview of the humoral immune system

The humoral immune response includes differentiation of mature B cells after (a) an initial interaction with Ag into (b) Ab-secreting plasma cells. (c) Antibodies act as effector molecules that neutralise and clear pathogens. (d) Visualisation of a mature B cell using a scanning electron microscope (Adapted from Janeway *et al.*, 2005).

The humoral immune system is the interaction of B cells with an Ag and subsequently their differentiation into Ab-secreting plasma cells or memory B cells (Figure 1.1). The humoral immune system acquired its name as it involves molecules that are present in the humours, or body fluids. Ab molecules function as effector molecules by recognising, neutralising and eliminating Ags (Figure 1.1). The removal of Ags involves the cross-linking of an Ag by multiple Abs, forming a large complex that can be phagocytosed or opsonised. Binding of an Ab to an Ag on a microorganism can also activate complement, a branch of the innate immune system that can lyse and destroy the invading microorganism. The binding of Abs to an Ag can also prevent their attachment to, and consequent invasion of host cells.

1.1.4 B-cell maturation and effector functions

As shown in Figure 1.2, B cell development commences in the bone marrow, where a lymphoid stem differentiates into the earliest cell distinctive in the B cell lineage, namely a progenitor B cell (pro-B cell). Pro-B cells proliferate within bone marrow and then differentiate into precursor B cells (pre-B cell), for which the microenvironment of bone-marrow stromal cells is required. Stromal cells secrete interleukin 7 (IL-7), which drives B cell maturation by down-regulating pre-B cell adhesion molecule expression to encourage their detachment from the stromal cells.

During pre-B cell maturation into immature B cells, immunoglobulin (Ig) gene rearrangement occurs. The process of Ig gene rearrangement allows B cells to code for and express one functional heavy-chain and light-chain variable region in the Ab molecule. This makes each individual B cell antigenically committed to a specific epitope. Immature B cells express the μ isotype of the Ig constant heavy chain gene. This bone marrow phase of B cell development therefore culminates in the production of B cells that express membrane-bound immunoglobulin M (IgM). At this stage, the immature B cells are not fully functional and interaction with an Ag will induce cell death or anergy rather than proliferation and differentiation. Maturation of B cells is signalled by the co-expression of the δ and μ heavy chain constant genes, leading to the production

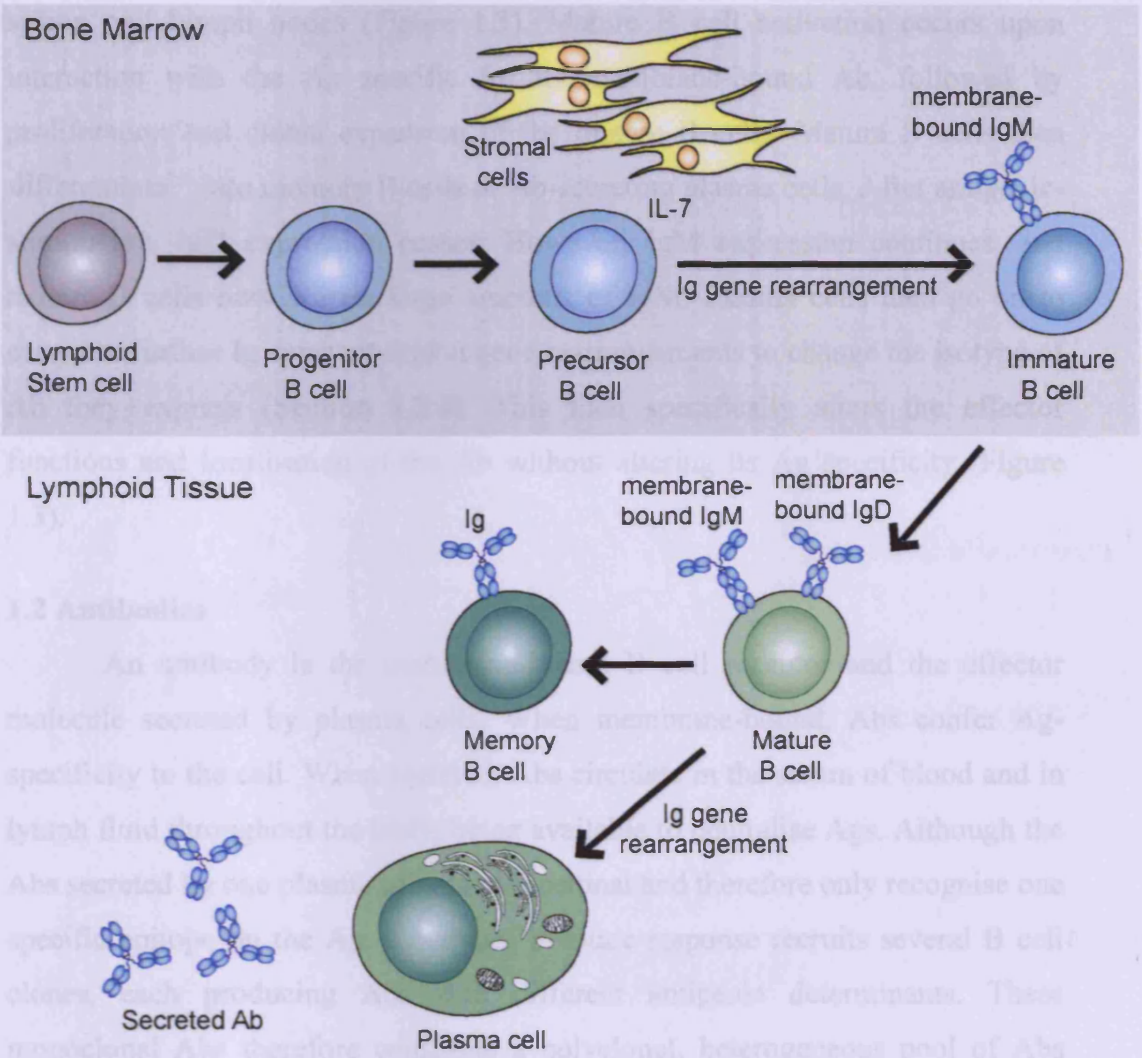

Figure 1.2 B cell maturation

Development and differentiation of a B cell from a non-committed lymphoid stem cell to either a memory B cell or an Ab-secreting plasma cell. Membrane-bound IgD and IgM are shown. Interleukin-7 (Il-7) is released by stromal cells in the bone marrow to drive pre-B cell maturation. If an immature B-cell encounters an Ag it will induce cell death or anergy.

of membrane-bound IgM and membrane-bound immunoglobulin D (IgD) (Figure 1.2). Each mature B cell expresses approximately 1.5×10^5 Ab molecules on its cell surface (Figure 1.1d), with specificity for a single Ag. Mature B cells circulate the body in the lymph fluid and blood to lymphoid organs such as the spleen and lymph nodes (Figure 1.3). Mature B cell activation occurs upon interaction with the Ag specific for its membrane-bound Ab, followed by proliferation and clonal expansion of the mature B cells. Mature B cells then differentiate into memory B cells or Ab-secreting plasma cells. After antigenic-stimulation, IgD expression ceases. However, IgM expression continues, and mature B cells now secrete large amounts of IgM. Plasma cells then go on to carry out further Ig constant-region gene rearrangements to change the isotype of Ab they express (Section 1.2.4). This then specifically alters the effector functions and localisation of the Ab without altering its Ag specificity (Figure 1.3).

1.2 Antibodies

An antibody is the membrane-bound B cell receptor and the effector molecule secreted by plasma cells. When membrane-bound, Abs confer Ag-specificity to the cell. When secreted, Abs circulate in the serum of blood and in lymph fluid throughout the body, being available to neutralise Ags. Although the Abs secreted by one plasma cell are monoclonal and therefore only recognise one specific epitope on the Ag, a humoral immune response recruits several B cell clones, each producing Abs with different antigenic determinants. These monoclonal Abs therefore constitute a polyclonal, heterogeneous pool of Abs that can potentially recognise a variety of epitopes on a pathogen.

1.2.1 General antibody structure

Antibodies have a generic common structure, consisting of four polypeptide chains; two identical immunoglobulin heavy chains (H) and two identical immunoglobulin light chains (L). Each L chain is generally bound to an H chain by a covalent disulphide bond, and also by non-covalent salt-bridges, hydrogen bonds and hydrophobic interactions, to form a heterodimer (H-L). Two identical heterodimer H chains are then covalently linked $(H\text{-}L)_2$. H and L chains each contain constant (C_H and C_L) and variable (V_H and V_L) regions (Figure 1.4).

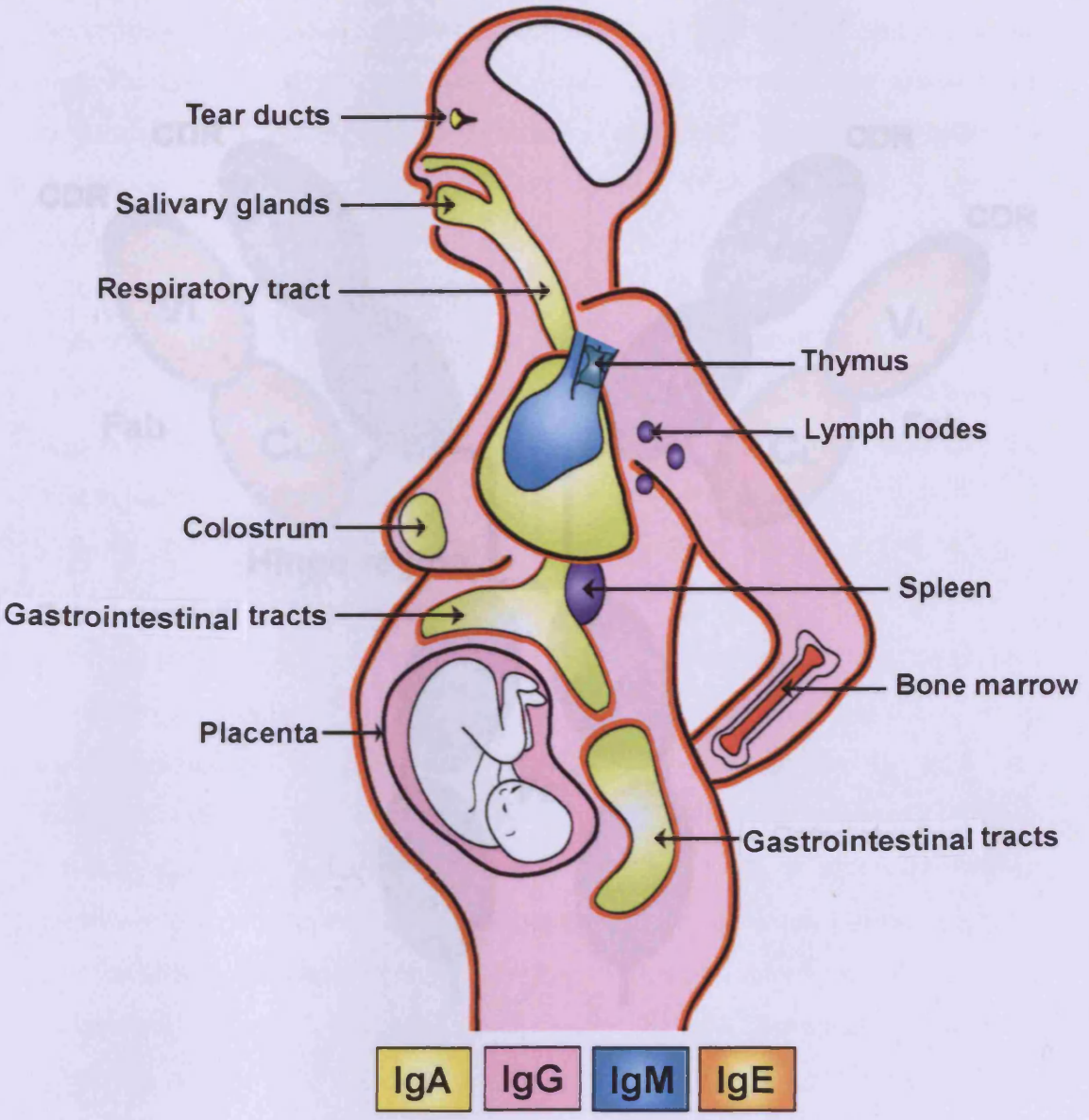

Figure 1.3 Distribution of effector sites of the different antibody classes throughout the body

IgA in its secretory form is the predominant Ab class present in mucosal secretions, including the gastrointestinal, respiratory and urogenital tracts, breast milk, saliva and tears. The main Ab classes in plasma are IgG and IgM, and in extracellular fluid are IgG and monomeric IgA. IgE is mainly associated with mast cells beneath epithelial surfaces. The neonate can receive IgG from the mother by transplacental transport. No antibodies are usually present in the brain.

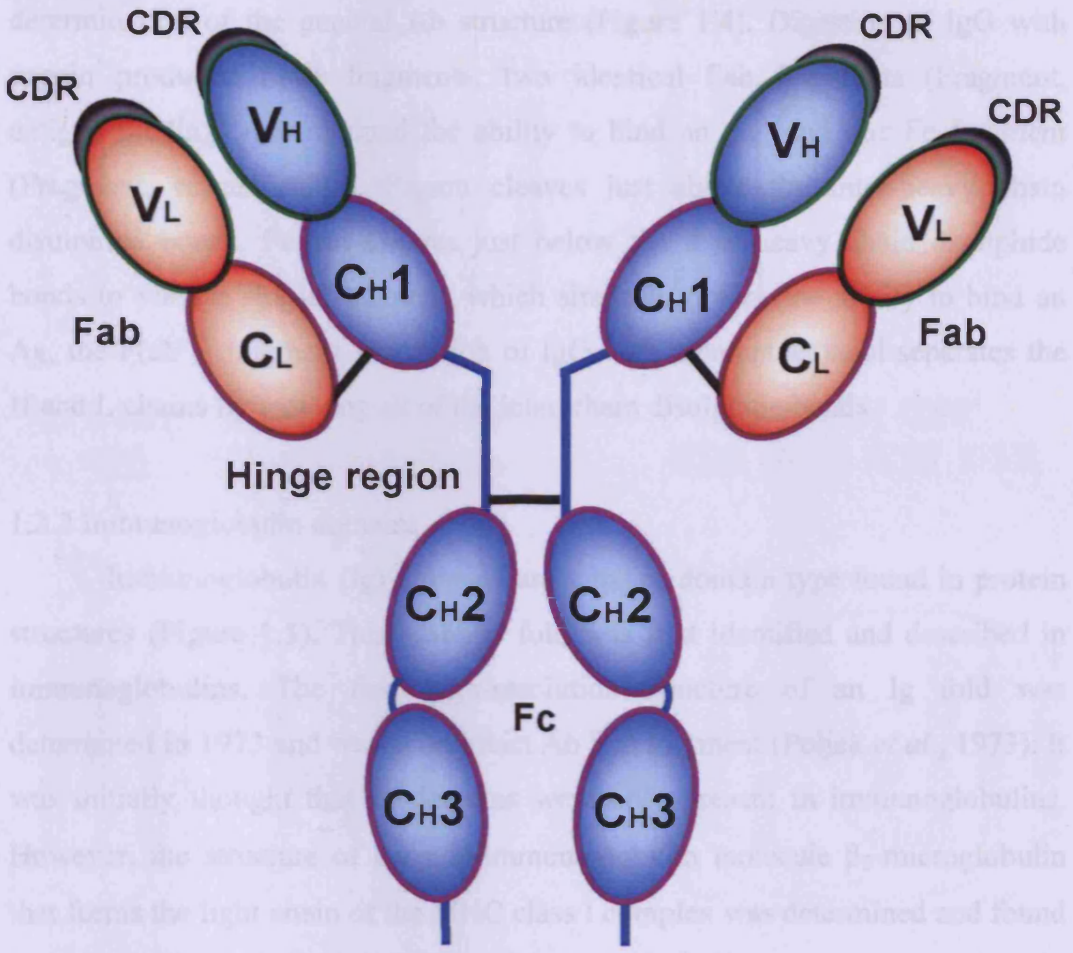

Figure 1.4 Generic antibody structure

Antibodies consist of two identical heavy chains (blue) and two identical light chains (red). The heavy and light chains are divided into variable (green outline) and constant (purple outline) regions. The heavy and light chain domains and the Fab and Fc regions are highlighted. The hinge and domain linkers are shown as blue lines. Inter-chain disulphide bonds are shown as black lines. Each Ig domain (oval) contains two intra-chain disulphide bonds (not shown). Complementarity determining regions (CDRs) are shown as black crescents.

The general structure of Abs was discovered by experimental observations of enzymatic cleavages. Immunoglobulin G (IgG) in serum was the first Ab class identified, and was subjected to experiments that allowed the determination of the general Ab structure (Figure 1.4). Digestion of IgG with papain produced three fragments; two identical Fab fragments (Fragment, antigen binding) that retained the ability to bind an Ag, and one Fc fragment (Fragment, crystallisable). Papain cleaves just above the inter-heavy chain disulphide bonds. Pepsin cleaves just below the inter-heavy chain disulphide bonds to yield a single fragment, which similarly retains the ability to bind an Ag, the $F(ab')_2$ fragment. Reduction of IgG with mercaptoethanol separates the H and L chains by reducing all of the inter-chain disulphide bonds.

1.2.2 Immunoglobulin domains

Immunoglobulin (Ig) domains are a major domain type found in protein structures (Figure 1.5). This globular fold was first identified and described in immunoglobulins. The first high-resolution structure of an Ig fold was determined in 1973 and was of an intact Ab Fab fragment (Poljak *et al.*, 1973). It was initially thought that Ig domains were only present in immunoglobulins. However, the structure of the non-immunoglobulin molecule β_2-microglobulin that forms the light chain of the MHC class I complex was determined and found to consist of a single Ig domain, homologous to that of a Ig constant-type domain (Peterson *et al.*, 1972). It therefore became apparent that the Ig fold has a greater significance in protein structures. The term 'Ig superfamily' is now used to describe proteins that contain Ig domains and the Ig fold has been described as one of the nine 'superfold' structures (Orengo *et al.*, 1994). A superfamily describes proteins whose corresponding genes are derived from a common primordial gene encoding the basic domain structure. There is a widespread occurrence of Ig domains in cell-surface receptors (Williams & Barclay, 1988), including Ab Fc receptors and the polymeric immunoglobulin receptor (pIgR). Also, cell adhesion molecules and muscle protein domains have been found to contain Ig domains (Smith & Xhe, 1997).

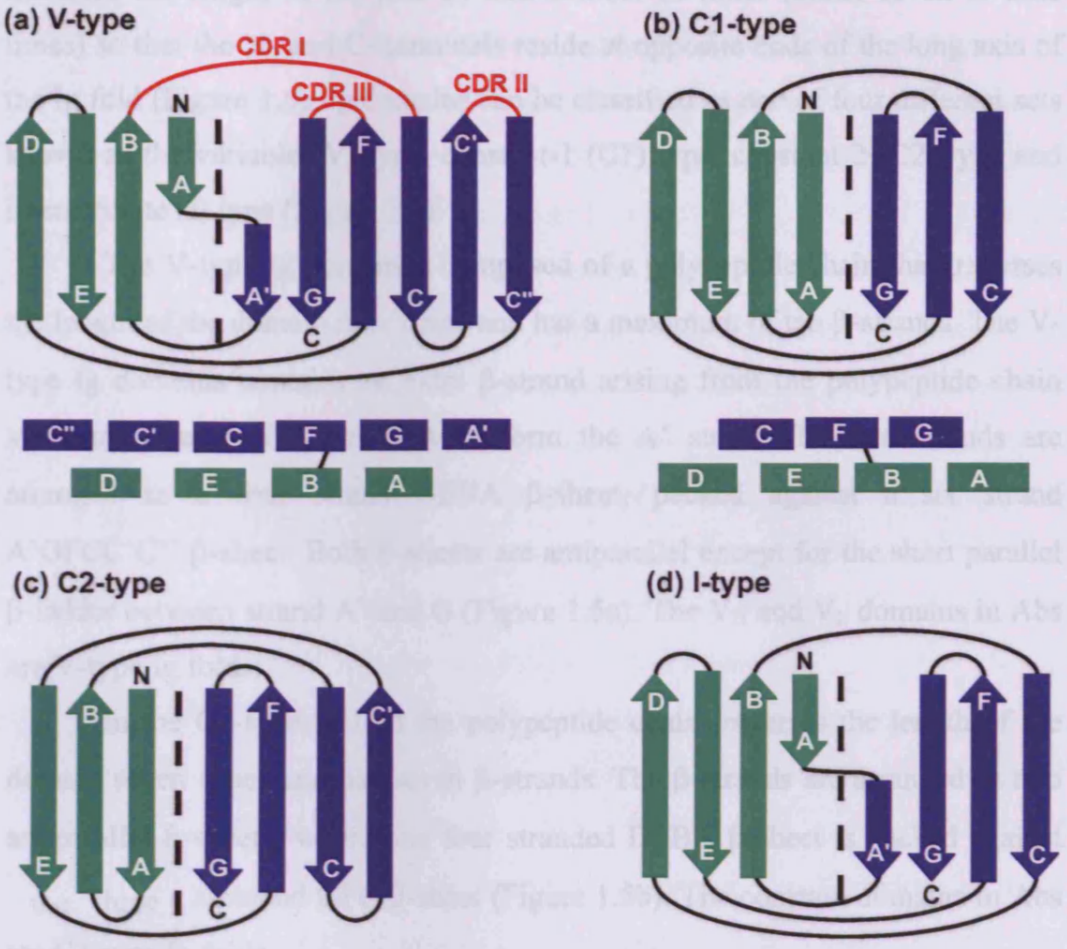

Figure 1.5 Two-dimensional topology diagrams for each of the four Ig folds

(a-d) An arrow represents the direction of the seven to ten β-strands (A, A', B, C, C', C'', D, E, F, G) and the direction from N to C terminus. The three loops forming the CDRs of the Ig V-type domain (red) are shown between β-strands B and C, C' and C'' and F and G. The β-strands viewed end-on as a β-sheet sandwich fold for the V and C1 type Ig folds are shown underneath. The intradomain disulphide bond in the Ig folds is shown with a straight black line.

Ig domains are all β-class protein folds that consist of two antiparallel β-sheets packed face-to-face to form a sandwich structure. The polypeptide chain traverses the length of the fold an odd number of times (either seven or nine times) so that the N- and C- terminals reside at opposite ends of the long axis of the Ig fold (Figure 1.5). Ig domains can be classified as one of four different sets known as the variable (V) type, constant-1 (C1) type, constant-2 (C2) type and intermediate (I) type (Figure 1.5).

The V-type Ig domain is composed of a polypeptide chain that traverses the length of the domain nine times and has a maximum of ten β-strands. The V-type Ig domains contains an extra β-strand arising from the polypeptide chain switching sheets after strand A to form the A' strand. The ten strands are arranged as a four strand DEBA β-sheet, packed against a six strand A'GFCC'C'' β-sheet. Both β-sheets are antiparallel except for the short parallel β-ladder between strand A' and G (Figure 1.5a). The V_H and V_L domains in Abs are V-type Ig folds.

In the C1-type Ig fold the polypeptide chain traverses the length of the domain seven times and has seven β-strands. The β-strands are arranged as two antiparallel β-sheets, where one four stranded DEBA β-sheet is packed against the three stranded GFC β-sheet (Figure 1.5b). The constant domains in Abs are C1-type Ig folds.

The C2-type Ig domain consists of a polypeptide chain traversing the length of the domain seven times. The C2-type fold is characterised by a peptide connection between the end of β-strand C and the beginning of β-strand E. This causes the β-strand D present in the C1-type fold to switch β-sheets and form a C' β-strand. The C2-type fold is consequently also called the switch (S)-type domain. Three β-strands are arranged as one EBA β-sheet packed against a four strand GFCC' β-sheet (Figure 1.5c).

The I-type of Ig fold has a polypeptide chain that traverses the domain length seven times. The I-type domain has eight β-strands, with the polypeptide chain switching sheets after A to also form A' similar to the V-type fold. The eight strands form one DEBA β-sheet with four stands packed against an A'GFC

β-sheet. Both β-sheets are antiparallel except for the short parallel β-ladder between strand A' and strand G (Figure 1.5d).

1.2.2.1 Ig domain common core and intradomain disulphide bonds

Even though there are significant structural disparities between the Ig domain types, a structural core is common to all the types of Ig domains (Bork *et al.*, 1994). All four Ig folds contain the four β-strands B, C, E and F (Figure 1.5). Strands B and E form an antiparallel β-ladder in the centre of one sheet that packs against and interacts with an antiparallel β-ladder formed by strands F and C in the centre of the other sheet. This structural core is stabilised by the aligned packing of mainly hydrophobic residue side chains, to form a hydrophobic core. The variation between the four Ig-fold types (and slight variation between domains within each Ig fold type) result from differences in the length and presence of the other strands, A, A', C, C', C'' and D, and the surrounding loop regions, which are present at the periphery of the β-sheets.

The V-type and C1-type of Ig folds contain a conserved intradomain disulphide bridge, usually formed between two cysteine residues in β-stands B and F (Lesk & Chothia, 1982; Figure 1.5a,b). It is thought that the presence of this disulphide bond increases the stability of the domain. However, in some domains these cysteine residues may be absent and in other domains the position or number of disulphide bonds is increased (Williams & Barclay, 1988; Bork *et al.*, 1994).

1.2.2.2 Complementarity Determining Regions

The differences between Ab molecule colonies that confer specific epitope recognition arise due to differences in the amino acid sequence in discrete regions of the V_H and V_L chains, called complementarity-determining regions (CDRs). CDRs are hypervariable regions that constitute the Ag-binding site. X-ray crystallography studies have established that the Ag-binding sites on an Ab are formed by three CDRs in the light chain (CDRI-CDRIII) and three corresponding CDRs (CDRI-CDRIII) in the heavy chain, all of which have high degree of sequence variability (Kabat *et al.*, 1976). The three CDRs are in loop regions at the same end of the β-barrel, connecting β-strands B-C, C'-C'' and F-

G (Figure 1.5a). The loop between the two extra β-strands in the V-type Ig domain, C' and C'', which is not present in the other classes of Ig fold, contains the hypervariable region CDRII (Figure 1.5a).

CDRs are regions of sequence hypervariability, yet within the same Ab class, very few sequence differences are seen throughout the rest of the Ab molecule. Therefore, the locations of CDRs are determined by comparing the sequences of light and heavy chains to determine those areas of significant variability. Plotting sequence variability as a function of residue number along a polypeptide chain for hundreds of different variable domains shows three large peaks corresponding to the hypervariable regions CDRI, CDRII and CDRIII (Wu *et al.*, 1970). CDRs have been defined as heavy chain residues 31-35, 50-65 and 95-102 and light chain residues 24-34, 50-56 and 89-97 (Figure 1.6).

Residues within CDRs are thought to have different roles, including (1) residues whose side-chains are in direct contact with the antigenic determinant and contribute to the binding-energy of the Ag-Ab interaction, (2) residues which fulfil a structural role and whose side-chains do not contact the Ag and (3) residues which are not in contact with the Ag but may exert a conformational influence on contacting residues and are therefore essential to the specificity (Kabat *et al.*, 1976).

1.2.3 Fab arm and antigen-binding site formation

The hypervariable CDRs from both the heavy chain V_H and light chain V_L have a close association to form an Ag-binding site. The Fab arms of an Ab are arranged in two globular regions where C_H1 associates with C_L, and V_H associates with V_L (Figure 1.4). The six CDRs are positioned together at the end of the variable domains and Fab arms meaning that they are free to interact with their specific Ag with little steric hindrance.

The C_H1 and C_L domains are paired through their DEBA four-stand β-sheet. The interface between the two domains contains close-packed hydrophobic residues. The β-sheets in the two constant domains are arranged at near-right angles to one another, burying the central hydrophobic residues. However, the V_H and V_L domains interact very differently. If the interaction between the two

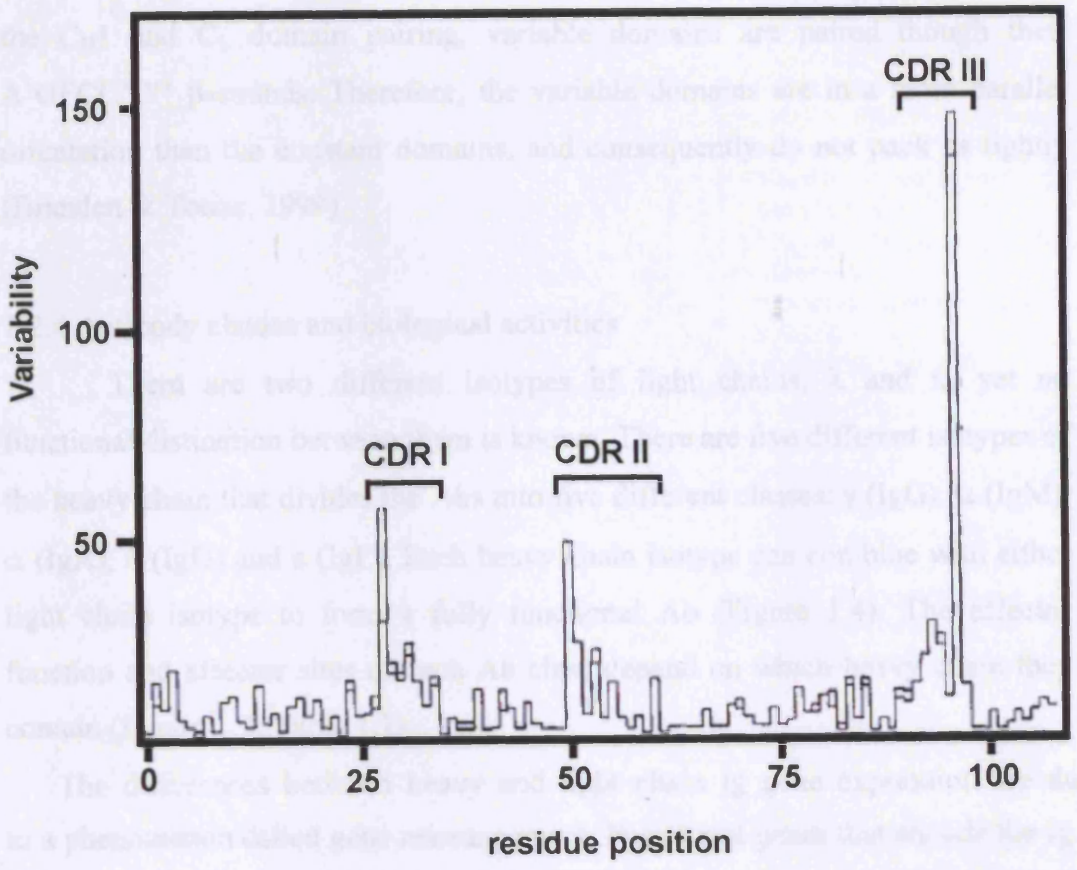

Figure 1.6 Location of the complementarity determining regions (CDRs) in variable immunoglobulin domains of antibody light chains

The regions of high sequence variability are highlighted and constitute CDRI (residues 24-34), CDRII (residues 50-56) and CDRIII (residues 89-97). The Y-axis variability is defined as the number of different amino acids at a given position divided by the frequency of the most frequently occurring amino acid at that position, where the frequency is calculated from the number of times the most common amino acid occurs divided by the total number of proteins studied. (Adapted from Wu *et al.*, 1970).

variable domains, V_H and V_L were similar to that between C_H1 and C_L where the four-stranded DEBA β-sheets interacted, it would mean that the loops containing the CDRs would be positioned far apart from one another as the majority of the CDRs are loops between the A'GFCC'C'' β-strands (Figure 1.5a). In contrast to the C_H1 and C_L domain pairing, variable domains are paired though their A'GFCC'C'' β-strands. Therefore, the variable domains are in a more parallel orientation than the constant domains, and consequently do not pack as tightly (Branden & Tooze, 1999).

1.2.4 Antibody classes and biological activities

There are two different isotypes of light chains, λ and κ, yet no functional distinction between them is known. There are five different isotypes of the heavy chain that divides the Abs into five different classes: γ (IgG), μ (IgM), α (IgA), δ (IgD) and ε (IgE). Each heavy chain isotype can combine with either light chain isotype to form a fully functional Ab (Figure 1.4). The effector function and effector sites of each Ab class depend on which heavy chain they contain (Figure 1.3; Table 1.1).

The differences between heavy and light chain Ig gene expression are due to a phenomenon called gene rearrangement. Functional genes that encode the Ig heavy and light chain are assembled at recombinational events at the DNA level. There are many genes encoding the light and heavy chains of Ab molecules. Light chain polypeptides arise from a short leader (L) exon, a variable (V) gene segment (λ or κ), a joining (J) gene segment and a light chain constant gene segment (λ or κ). Light chain Ig genes are rearranged to form a polypeptide consisting of VJC (of the λ or κ isotype). Like the light chain, heavy chains possess these gene segments as well as a diversity (D) gene segment, yet their constant gene segments are α, δ, ε, γ or μ. Heavy chain gene rearrangement produces polypeptides consisting of $V_HD_HJ_HC_H$. The C_H regions code for the heavy chains of the Ab, and are present as exons separated by introns in the order 5'-μ-δ-γ3-γ1-α1-γ2-γ4-ε-α2-3'. An analysis of functional light and heavy chain V, D and J gene segments in the germ-line DNA of one human individual revealed 51 V_H, 25 D_H, 6 J_H, 40 $V_κ$, 5 $J_κ$, 31 $V_λ$ and 4 $J_λ$ segments. This

Antibody Class		Mw (kDa)	Heavy Chain	Serum levels (mg/ml)	*In Vivo* half life (days)	Biological activities
IgA	IgA1	150-600	α1	3	6	▪ Main Ab present in external secretions (mucus, saliva, tears) ▪ Present in breast milk for immune system of neonate ▪ First line of defence in mucosal immunity ▪ When polymeric in secretions can cross-link large Ag ▪ Activates complement pathway
	IgA2	150-600	α2	0.5	6	▪ Mucosal immunity ▪ Present in secretions (mucus, saliva, tears) ▪ Present in breast milk
IgG	IgG1	150	γ1	9	23	▪ Binds to phagocyte Fc receptors and mediates opsonisation ++ ▪ Activates complement pathway + ▪ Crosses the placenta +
	IgG2	150	γ2	3	23	▪ Binds to phagocyte Fc receptors and mediates opsonisation +/- ▪ Activates complement pathway +/- ▪ Crosses the placenta +/-
	IgG3	150	γ3	1	8	▪ Activates classical complement pathway ++ ▪ Binds to phagocyte Fc receptors and mediates opsonisation ++ ▪ Crosses the placenta +
	IgG4	150	γ4	0.5	23	▪ Crosses the placenta + ▪ Binds to phagocyte Fc receptors +
IgM		180-900	μ	1.5	5	▪ First Ab class produced in a primary immune response ▪ First Ab class synthesised by the neonate ▪ Present on mature B cell membranes ▪ High valency when pentameric to bind Ag with repeating epitopes, e.g. viral particles ▪ Activates complement pathway +++ ▪ Present in external secretions with accessory role as secretory Ab
IgD		150	δ	0.03	3	▪ Present on mature B cell membranes
IgE		190	ε	0.0003	2.5	▪ Mediates hypersensitivity reactions (allergies) ▪ Induces basophil and mast cell degranulation

Table 1.1 Antibody bioactivities

(Adapted from Goldsby *et al.*, 2002).

mechanism of Ig gene rearrangement therefore generates a minimum of 1.5×10^6 combinations and subsequently a huge Ab repertoire.

B cells can undergo further gene rearrangement during their lifetime, in which their $V_HD_HJ_H$ unit can combine with different C_H gene segments, stimulated by specific cytokine signals that can dictate the Ab isotype to which the B cell switches. This process is known as class switching. This allows the immune system to control which Ab class is produced by the B cells, creating specific effector responses in the correct area of the human body (Figure 1.3; Table 1.1).

1.2.4.1 Immunoglobulin G

IgG is the most abundant Ab class present in human serum, present at approximately 12.3 mg/ml, which constitutes around 80% of the total serum Ab content (Table 1.2). It is monomeric with a molecular mass of 150 kDa. It consists of two γ heavy chains bound to two κ or λ light chains. The heavy chain of IgG has three constant domains, C_H1, C_H2 and C_H3, similar to the general Ab structure shown in Figure 1.4. Four subclasses of IgG exist; IgG1, IgG2, IgG3 and IgG4. The different subclasses of IgG all possess between 90-95% sequence similarity, with the main sequence and structural disparity present in the hinge region connecting the Fc region to the Fab fragments, and the number and position of inter-heavy chain disulphide bonds. These structural differences give the IgG subclasses varying biological potency in activating complement, their ability to cross the placenta and their affinity of binding to their Fc receptor on phagocytes and triggering opsonisation (Figure 1.3; Table 1.1)

1.2.4.2 Immunoglobulin M

IgM is the first Ab class to be expressed on immature B cells and also the first Ab class secreted in a primary immune response (Figure 1.2). IgM accounts for approximately 5-10% of the total serum immunoglobulins, and is at a serum concentration of ~1.5 mg/ml. Monomeric IgM has a molecular mass of around 180 kDa and is membrane-bound. The heavy chain of IgM contains an extra constant domain to the generic Ab structure that is designated C_H4 (Figure 1.4; Figure 1.7). When secreted, IgM is pentameric with a molecular mass of

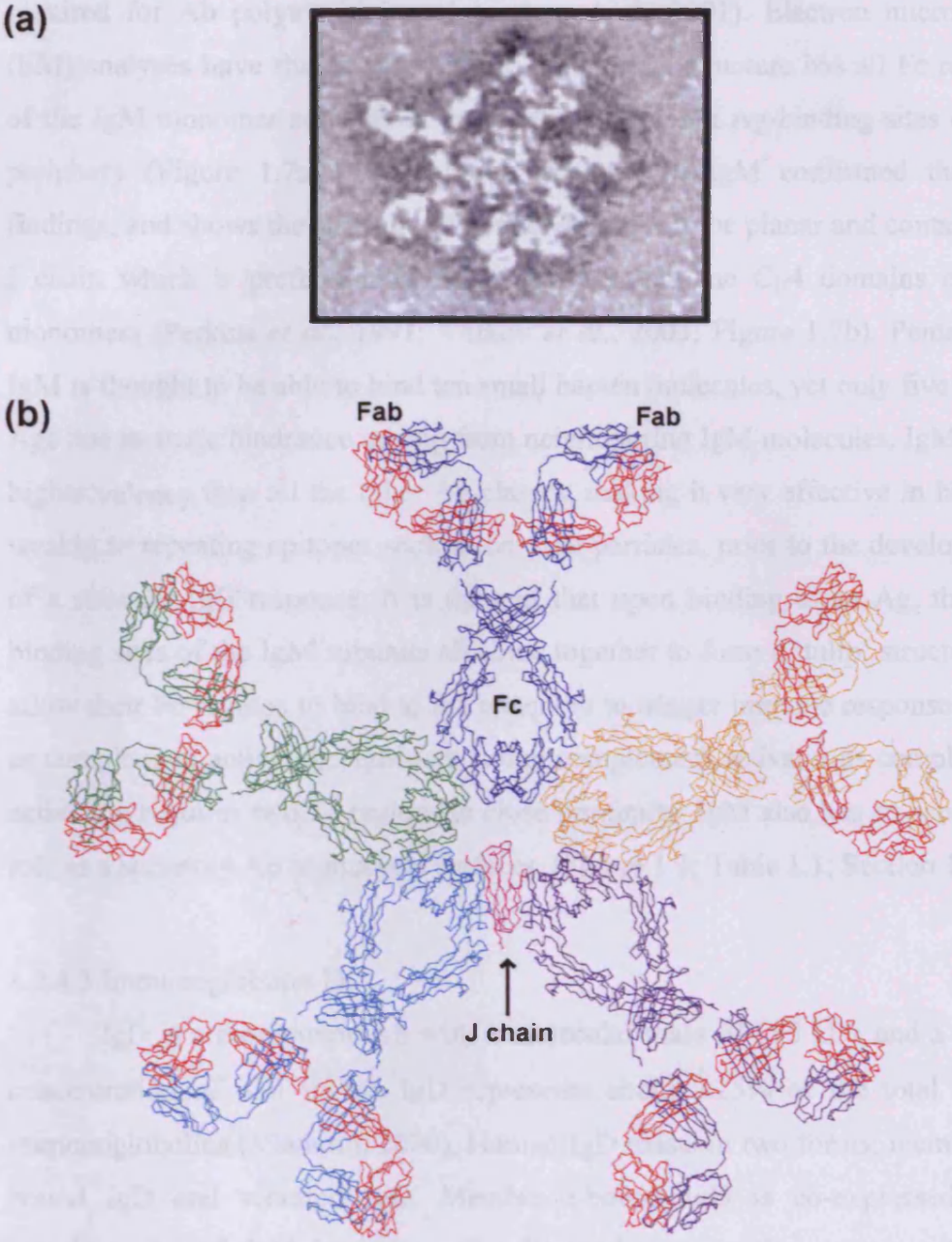

Figure 1.7 Electron micrograph and solution structure of pentameric IgM
(a) Visualisation of pentameric IgM molecules by scanning electron microscopy (Adapted from Janeway *et al.*, 2005). (b) Solution structure of pentameric IgM (PDB code 2RCJ; Perkins *et al.*, 1991). The light chains of the five IgM monomers are shown in red, the heavy chains are shown in dark blue, green, light blue, purple and orange and the J chain is shown in pink.

around 900 kDa (Figure 1.7). It also contains an 18 kDa joining (J) chain that is required for Ab polymerisation (Johansen *et al.*, 2001). Electron microscopy (EM) analyses have shown that the IgM pentameric structure has all Fc regions of the IgM monomer subunits in the centre and the ten Ag-binding sites on the periphery (Figure 1.7a). The solution structure of IgM confirmed the EM findings, and shows the structure of pentameric IgM to be planar and contain one J chain which is preferentially positioned between the C_H4 domains of two monomers (Perkins *et al.*, 1991; Volkov *et al.*, 2003; Figure 1.7b). Pentameric IgM is thought to be able to bind ten small hapten molecules, yet only five larger Ags due to steric hindrance arising from neighbouring IgM molecules. IgM has a higher valency than all the other Ab classes, making it very effective in binding weakly to repeating epitopes such as on viral particles, prior to the development of a stronger IgG response. It is thought that upon binding to an Ag, the Ag-binding sites of the IgM subunits all come together to form a 'tulip' structure, to allow their Fc regions to bind to the receptors to trigger immune responses such as complement activation. IgM is a strong complement activator as complement activation requires two Fc regions in close proximity. IgM also has an accessory role as a secretory Ab in mucosal surfaces. (Figure 1.3; Table 1.1; Section 1.4.3)

1.2.4.3 Immunoglobulin D

IgD is a monomeric Ab with a molecular mass of 185 kDa and a serum concentration of ~30 μg/ml. IgD represents about 0.25% of the total serum immunoglobulins (Vladutiu, 2000). Human IgD exists in two forms; membrane-bound IgD and secreted IgD. Membrane-bound IgD is co-expressed with membrane-bound IgM by mature B cells, and together they act as antigenic receptors on the surface of the lymphocytes (Figure 1.2). IgD is most abundant in its membrane-bound form and little is known about the function of secreted IgD (Geisberger *et al.*, 2006). The majority of IgD-expressing plasma cells have been identified in the upper respiratory tract, particularly in the nasal passages and the tonsils, and few have been found in the bone marrow (Preud'homme, 2000)

IgD has three constant heavy chain domains and the longest hinge of all Ab classes, with 64-residues connecting the Fab and Fc regions. Secreted IgD has been shown to be more sensitive to proteolytic enzymes than IgM, and this

may be due to the long hinge region (Figure 1.4; Geisberger *et al.*, 2006). The solution structure of serum myeloma IgD has been determined, which has shown IgD to possess a flexible T-shape structure, unlike the Y-shape seen for IgG (Sun *et al.*, 2005). IgD contains a single intra-heavy chain disulphide bond, close to the C-terminus, permitting IgD to have a greater degree of freedom for Ag-binding than the other Ab isotypes (Table 1.1).

1.2.4.4 Immunoglobulin E

Immunoglobulin E (IgE) is monomeric with a molecular mass of 190 kDa and it is present in very low concentrations in serum (around 0.3 μg/ml). The heavy chain of IgE contains an extra constant domain compared to the generic Ab structure that is designated C_H4 (Figure 1.4). IgE mediates hypersensitivity reactions that are responsible for allergic reactions and anaphylactic shock. Cross-linking of Ag-bound IgE binding to their to Fc receptors on the membranes of basophils and tissue mast cells induces these cells to exocytose their granules in a process called degranulation. Pharmacologically active mediators are released to give rise to allergic manifestations. IgE is therefore mainly present below the surface of the epidermal tissue (Figure 1.3; Table 1.1).

1.3 Immunoglobulin A

Immunoglobulin A (IgA) is by far the most abundant antibody present in the human body. The daily production of IgA is estimated to be of the order of 65 mg/kg, far exceeding the production of all the other antibody classes combined (Kerr, 1990). In its secretory form, IgA is the principle antibody on mucosal surfaces (Figure 1.3). In addition, IgA is the second most prevalent antibody in circulation in human serum. IgA is present in very different environments in the human body (Table 1.1). This is reflected in the unique heterogeneity in the molecular forms of IgA and characteristic distribution in the various body fluids (Table 1.2).

IgA exists as two subclasses, IgA1 and IgA2 (Figure 1.8), the latter of which is found in at least two allotypic forms, IgA2m(1) and IgA2m(2). IgA2n represents a possible third allotype of IgA2 (Chintalacharuvu *et al.*, 1994). Serum IgA is predominantly monomeric IgA1, with a concentration of about 2-3 mg/ml and low percentage of polymeric forms. In secretions, the IgA1 and IgA2

Human Fluid	IgG (mg/ml)	IgA (mg/ml)	pIgA (%)	Ratio of IgA1/IgA2
Serum	12.3	3.28	13	89:11
Colostrum	0.1	12.34	96	65:35
Saliva	0.05	0.30	96	63:37
Jejunal	0.34	0.28	95	70:30
Colonic	0.86	0.83	n.d.	35:65
Hepatic bile	0.18	0.11	65	74:26
Nasal	0.06	0.26	n.d.	95:5
Bronchial	0.02	n.d.	82	67:33

Table 1.2 Concentrations of IgA in serum and human secretions and comparison to that of IgG

Those concentrations of IgA that are not detectable are denoted n.d. (Adapted from Kerr, 1990).

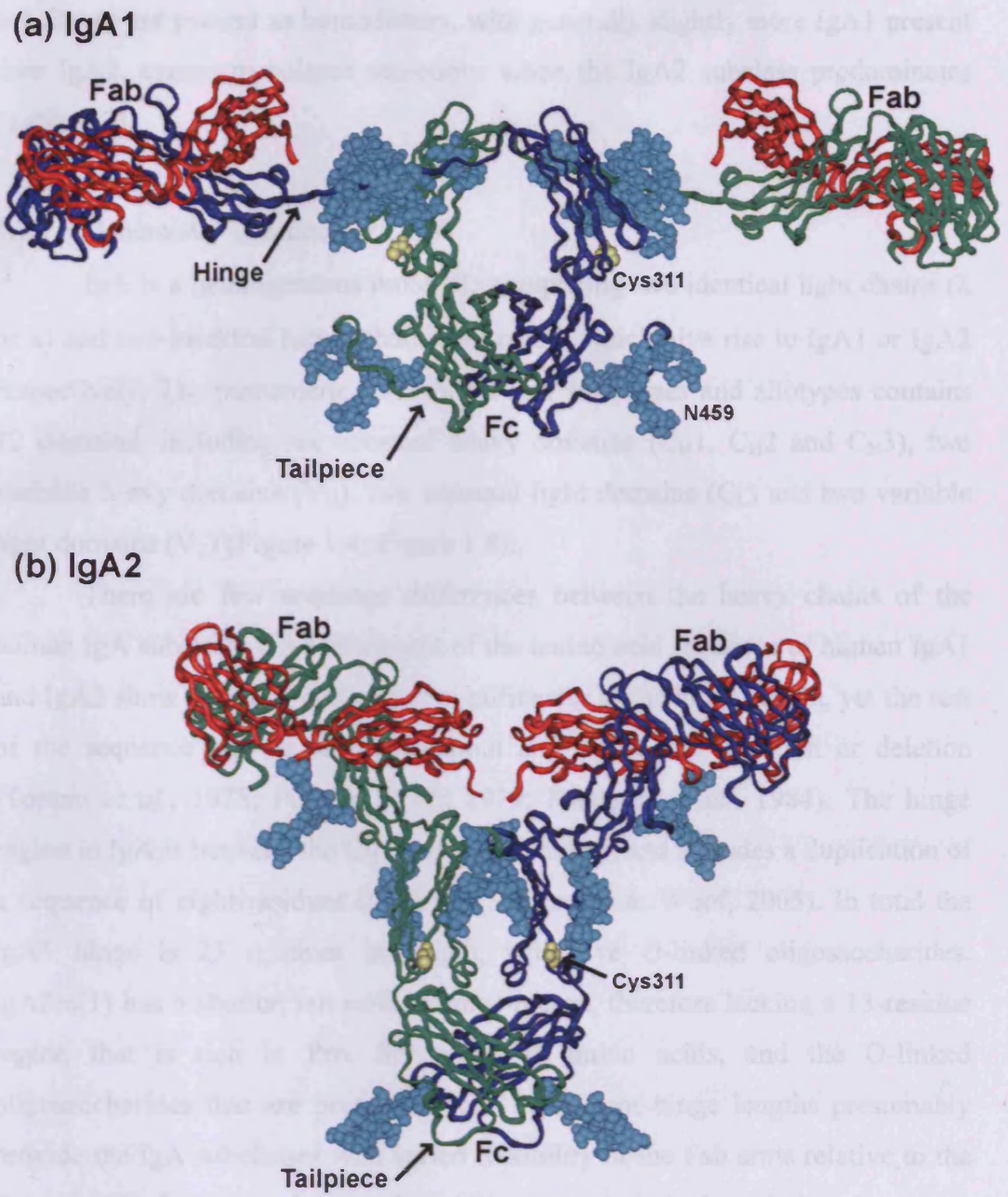

Figure 1.8 Solution structures of IgA1 and IgA2m(1)

The best-fit solution structures of (a) IgA1 (PDB 1IGA; Boehm *et al.*, 1999) and (b) IgA2m(1) (PDB 1R70; Furtado *et al.*, 2004) are shown. The antibody heavy chains (blue and green) and light chains (red), N-linked and O-linked carbohydrates (cyan) are shown. The Cys311 that covalently binds to SC to form SIgA (yellow) are highlighted in both structures.

subclasses are present as homodimers, with generally slightly more IgA1 present than IgA2, except in colonic secretions when the IgA2 subclass predominates (Table 1.2).

1.3.1 Biochemistry of monomeric IgA

IgA is a heterogeneous molecule comprising two identical light chains (λ or κ) and two identical heavy chains, α1 or α2, which give rise to IgA1 or IgA2 respectively. The monomeric form of all IgA subclasses and allotypes contains 12 domains, including six constant heavy domains (C_H1, C_H2 and C_H3), two variable heavy domains (V_H), two constant light domains (C_L) and two variable light domains (V_L) (Figure 1.4; Figure 1.8).

There are few sequence differences between the heavy chains of the human IgA subclasses. An alignment of the amino acid sequence of human IgA1 and IgA2 show that they differ most significantly in the hinge region, yet the rest of the sequence can be aligned without any significant insertion or deletion (Torano *et al.*, 1978; Putnam *et al.*, 1979; Flanagan *et al.*, 1984). The hinge region in IgA is between the C_H1 and C_H2 domains, and includes a duplication of a sequence of eight-residues (PSTPPTPS) (Senior & Woof, 2005). In total the IgA1 hinge is 23 residues in length, with five O-linked oligosaccharides. IgA2m(1) has a shorter, ten residue hinge region, therefore lacking a 13-residue region that is rich in Pro, Ser and Thr amino acids, and the O-linked oligosaccharides that are present in IgA1. Different hinge lengths presumably provide the IgA subclasses with varied flexibility of the Fab arms relative to the Fc region, antigenic-reach and vulnerability to proteolytic degradation.

The best characterised allotypic forms of human IgA2 are IgA2m(1) and IgA2m(2). The IgA2m(1) allotype appears to be a hybrid gene possibly arising from a recombination or gene conversion involving IgA1 and IgA2m(2), thus the C_H3 domain of IgA2m(1) is identical to that of IgA1, whereas the C_H1 and C_H2 domains are related to IgA2m(2) (Flanagan *et al.*, 1984). A major structural difference between IgA1 and IgA2 is the arrangement of inter-chain disulfide bridges. In IgA1, the light chains covalently bind to the heavy chains via Cys214 in the C_L domain and Cys133 in the C_H1 domain. However, there are no disulphide bonds present to link the heavy and light chains in IgA2m(1) (Grey *et*

al., 1968). This lack of inter-chain disulphide bonds is due to the presence of an asparagine residue instead of the Cys133 in its C_H1 domain. Instead, the two IgA2m(1) light chains bind via Cys214. Therfore, the IgA2m(1) light and heavy chain interact by non-covalent interactions (van Loghem & Biewenga, 1983; Furtado *et al.*, 2004).

Both IgA subclasses and IgA2 allotypes possess an additional 18-residue C-terminal tail after the C_H3 domain, termed the tailpiece (Figure 1.8). The tailpiece is involved in the polymerisation of the IgA monomers (Mestecky *et al.*, 1974). Cys471 in the tailpiece of one heavy chain in each monomer covalently binds to the J chain. The deletion of the tailpiece and mutation of Cys471 prevents IgA1 dimer assembly (Atkin *et al.*, 1996).

IgA sequences from many species have been determined. In mouse serum, only one subclass has been identified that is most similar to that of human IgA2, and exists only in polymeric forms (Kerr, 1990). In rabbits, 13 IgA subclasses have been identified with between 72% and 85% sequence similarity to each other, particularly in the C_H2 and C_H3 domains (Burnett *et al.*, 1989). The human and mouse IgA hinge regions contain no cysteine residues, whereas the rabbit IgA isoforms contain between one and five cysteine residues. The 13 rabbit IgA hinge sequences vary extensively, which may offer greater mucosal protection through differing susceptibility to proteases (Burnette *et al.*, 1989). Homologues of human IgA1 are only found in higher apes, and sequence comparisons show significant sequence hypervariability in the hinge region. IgA is a secretory Ab and IgA1 may be subject to positive selection, driven by its susceptibility to proteases in the mucosal environments (Sumiyama, *et al.*, 2002). When secreted, IgA1 is known to be susceptible to protease activity, rendering the Ab inactive by cleavage of its hinge region (Senior & Woof, 2005).

Electron micrographs of monomeric IgA show it to possess a Y-shaped structure, similar to that seen by IgG (Dourmashkin *et al.*, 1971). The solution structure of monomeric forms of IgA1 and IgA2m(1) have been determined (Figure 1.8; Boehm *et al.*, 1999; Furtado *et al.*, 2004). Although both are predominantly T-shaped, the structural differences between IgA1 and IgA2 due to the longer hinge region in IgA1 are evident (Figure 1.8). IgA1 has a more extended conformation than IgA2, which is thought to give IgA1 a greater ability to intercept and bind to Ags. The variation in distance between the Fab CDR

regions between IgA1 and IgA2 suggests they recognise different antigenic-surfaces on pathogens. The flexibility of the IgA1 hinge is thought to be reduced due to the presence of five O-linked oligosaccharides in each heavy chain of IgA1. The high percentage of proline residues and likewise the lack of glycine residues in both the IgA1 and IgA2 hinges are thought to decrease their hinge flexibility.

IgA is heavily glycosylated, and its carbohydrate content accounts for approximately 9% of the total mass of IgA, which is thought to contribute to it being more negatively charged than IgG (Boehm *et al.*, 1999; van Egmond *et al.*, 2001; Heremans, 1974). The carbohydrates residues may be involved in the protection of IgA, especially when in its secretory form in the harsh acidic and proteolytic environment of the mucosa. Within the IgA isotypes, IgA2 is more highly N-glycosylated than IgA1 (Tomana *et al.*, 1976). The IgA1 heavy chain contains two N-linked glycosylation sites at Asn263 within the C_H2 domain and at Asn459 within the tailpiece. IgA2m(1) contains two additional N-linked sites at positions Asn166 in the C_H1 domain and Asn337 in the C_H2 domain. IgA2m(2) contains all these sites plus one more at position Asn211 in the C_H2 domain. IgA1 alone possesses the five O-linked glycosylation sites in the hinge region. The O-linked glycosylation residues are Thr225, Thr228, Ser230, Ser232 and Thr236 in the hinge region (Mattu *et al.*, 1998). The presence of these O-linked carbohydrates offers protection to the longer IgA1 hinge, which is more vulnerable to proteolysis.

1.3.1.1 N- and O-linked carbohydrates

N-linked and O-linked glycosylation differ in several ways. N-linked carbohydrates are added as large 14-residue $GlcNAc_2Man_9Glc_3$ precursors to proteins co-translationally in the endoplasmic reticulum prior to folding. O-linked carbohydrates are assembled sequentially onto folded proteins in the Golgi apparatus, where N-linked carbohydrates may continue to be enzymatically-processed to their mature form (Gala & Morrison, 2002). O-linked carbohydrates are often shorter with between one and four residues. N-linked carbohydrates bind to the nitrogen in the R-group of asparagine, with the sequence Asn-X-Ser or Asn-X-Thr, where X is any amino acid except proline. O-linked carbohydrates

are added to serine or threonine residues, where an N-acetylgalactosamine (GalNAc) monosaccharide is linked to the amino acids' hydroxyl group.

1.3.2 Overview of the interaction of IgA with other proteins

Human serum albumin (HSA), α1-antitrypsin and α1-microglobulin (also known as Protein HC) can form complexes with IgA (Grubb *et al.*, 1983; Calero *et al.*, 1994). HSA binds to the IgA base of the Fc region but not to the Fab region (Almogren *et al.*, 2006). The level of these complexes in normal human serum is minute. However in some diseases the expression of these complexes is markedly increased (Vaerman, 1971). For example, high levels of IgA-HSA complex are found in some IgA myeloma cases (Mestecky *et al.*, 1977) and IgA-α1-antitrypsin complex is associated with the pathogenesis of rheumatoid arthritis (Stanworth *et al.*, 1985). Some reports have shown that IgA-α1-antitrypsin complex levels increase in the serum and synovial fluid of rheumatoid arthritis patients (Mattey *et al.*, 1997; Scott *et al.*, 1998).

IgA is considered a poor activator of complement in comparison to IgG and IgM. IgA does not activate complement by the classical pathway and does not bind C1q. IgA can activate the alternative pathway of complement, therefore mediating bacterial death via opsonisation or cell lysis. All isoforms of IgA are thought to be able to activate the alternative pathway, and an intact $F(ab')_2$ fragment is necessary (Hiemstra *et al.*, 1988). However, IgA activation of the alternative pathway remains a controversial topic, as in those studies IgA has to be either immobilised or aggregated, and the activation has not resulted in the formation of the membrane-attack complex. More recently, it has been shown that polymeric IgA, upon binding mannan-binding lectin (MBL) can activate the lectin pathway (Roos *et al.*, 2001). The role of IgA-complement activation *in vivo* remains unclear. With limited Ag present, IgA would actually inhibit complement activation by competing with IgG or IgM and blocking their binding to the Ag. If IgA does not activate complement or prevents complement activation, this may be important in preventing harmful inflammatory responses of complement system components. This is especially true in mucosal immunity, as IgA is in greater abundance in mucosal secretions and pro-inflammatory reactions may damage the mucosal tissue.

1.3.3 Dimeric IgA

Dimeric IgA (dIgA) is formed through the binding of Cys471 in the tailpiece of two IgA monomers to Cys15 and Cys68 of the J chain (Figure 1.9; Figure 1.10; Mestecky *et al.*, 1974; Garcia-Pardo, 1981). Electron micrographs of dIgA depict the two IgA monomers arranged with both Fc regions together, in a double Y-shape, with some flexibility (Figure 1.9; Svehag & Bloth, 1970; Dourmashkin *et al.*, 1971; Munn *et al.*, 1971). The slight variations in dIgA structure seen in these electron micrographs may be due to flexibility or the harsh staining techniques used in this technique (Section 2.1.1).

IgA is implicated in numerous diseases (Section 1.7), and in its dimeric form it is involved in the pathology of IgA nephropathy (IgAN). Serum dIgA1 levels are elevated in IgAN and there is deposition of IgA1 in the kidney glomerular mesangium (Section 1.7.2; Barratt *et al.*, 2007).

1.3.3.1 Joining chain

Despite the similarity of the overall structure of antibodies, their molecular masses vary dramatically from 150 kDa to 950 kDa, due to the ability of Abs to polymerise. The formation of dimers and higher oligomers is attributed to covalent binding of the Ab complex to the J chain. The J chain was first identified in 1970 (Halpern & Koshland, 1970) and was named after its ability to join two IgA monomers together to form dIgA.

The J chain is a 137-residue, 18 kDa glycoprotein with an acidic character due to its high glutamic and aspartic acid content. J chain contains one N-linked glycosylation site at Asn48 that accounts for approximately 12.5% of its molecular mass. Only one J chain is present in each polymeric Ab macromolecule independent of the size of the polymer (Mestecky *et al.*, 1974). One identical J chain is present in dIgA and pentameric IgM. The J chain dimerises IgA via its Cys15 and Cys68 covalently linking with two Cys471 residues in a tailpiece in each of the two IgA monomers (Figure 1.10). The carbohydrate moiety in J chain is also thought to be involved in its interaction with IgA (Krugman *et al.*, 1997). It was once thought that through its high cysteine content of eight residues, the J chain had enough free sulphyl-groups to bind to each heavy chain in each IgA or IgM in the complex, meaning that J

Figure 1.9 Electron micrographs of dimeric IgA and secretory IgA

(a) EM shows the structure of secretory IgA (SIgA) as a variable Y-shaped molecule and dIgA as a double Y-shape, with two IgA monomers bound at their Fc region. One dIgA Fab arm is arrowed and the scale bar represents 10 nm (Svehag & Bloth, 1970). (b) EM (×800,000) shows the dIgA structure as a double Y-shape. The scale bar represents 10 nm (Dourmashkin *et al.*, 1971). (c) Using a different EM staining technique as in (a), EM (×245,000) suggests SIgA has a similar arrangement of the two IgA monomers as in dIgA, unlike the images shown in (a) (Bloth & Svehag, 1971). (d) dIgA from (i) humans and (ii) mouse have similar overall double-Y structures. The scale bar represents 20 nm (Munn *et al.*, 1971). In (a) and (c) the SIgA preparations were obtained from colostrum and in (a-d) dIgA preparations were purified from human myeloma serum.

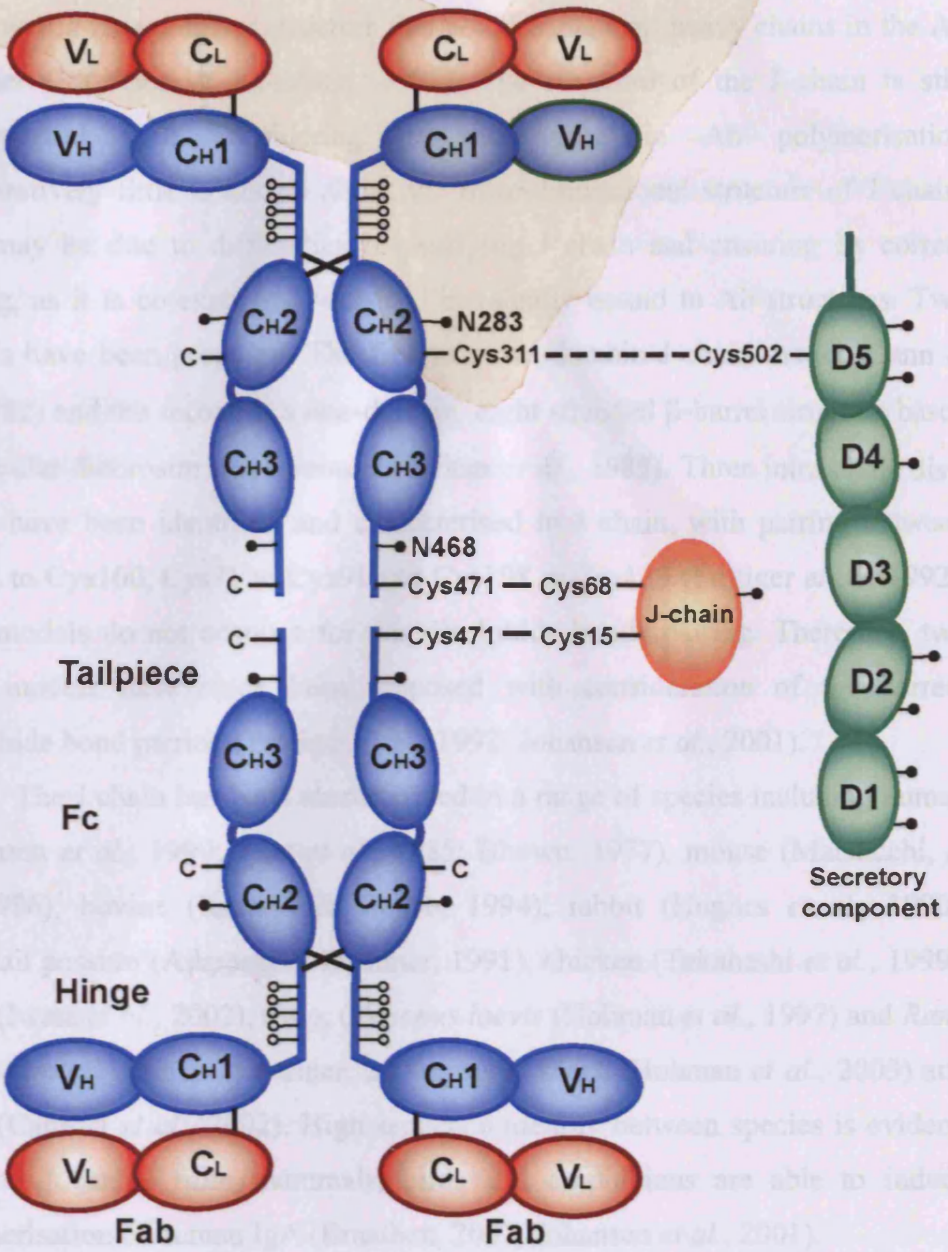

Figure 1.10 Schematic structure of dimeric/secretory IgA

dIgA is formed when two IgA monomers dimerise via covalently binding to J chain. When secretory component (SC) is covalently bound to dIgA (dashed line), SIgA is formed. Two IgA monomers are shown, with heavy chains (blue), light chains (red), O-linked carbohydrates (open circles), N-linked carbohydrates (filled circles) and inter-chain disulphide bonds (X) highlighted. The cysteine residues involved in disulphide bonds formation between IgA – J chain and IgA – SC are shown.

chain would have a linear structure that could contact all heavy chains in the Ab polymer (Morrison & Koshland, 1972). The structure of the J chain is still controversial, and considering its importance in Ab polymerisation comparatively little is known about the three-dimensional structure of J chain. This may be due to difficulties in purifying J chain and ensuring its correct folding, as it is co-expressed with and covalently bound to Ab structures. Two models have been proposed. The first is a two-domain J chain model (Cann *et al.*, 1982) and the second is a one-domain, eight stranded β-barrel structure based on circular-dichrosim measurements (Zikan *et al.*, 1985). Three intra-chain disulphide bonds have been identified and characterised in J chain, with pairing between Cys12 to Cys100, Cys71 to Cys91 and Cys198 to Cys133 (Frutiger *et al.*, 1992). Both models do not account for this disulphide bonds pairing. Therefore, two more models have since been proposed with consideration of the correct disulphide bond pairing (Frutiger *et al.*, 1992; Johansen *et al.*, 2001).

The J chain has been characterized in a range of species including human (Johansen *et al.*, 1999; Max *et al.*, 1985; Bhown, 1977), mouse (Matsuuchi, *et al.*, 1986), bovine (Kulseth & Rogne, 1994), rabbit (Hughes *et al.*, 1990), brushtail possum (Adamski & Demmer, 1991), chicken (Takahashi *et al.*, 1999), turtle (Iwata *et al.*, 2002), Frog, (*Xenopus laevis* (Hohman *et al.*, 1997) and *Rana catesbeiana* (Hohman & Steiner, 2000), nurse shark (Hohman *et al.*, 2003) and skate (Cannon *et al.*, 2002). High sequence identity between species is evident, and the J chains from mammals, birds and amphibians are able to induce polymerisation of human IgA (Braathen, 2007; Johansen *et al.*, 2001).

The J chain not only permits the polymerisation of IgA, it is also essential in the binding interaction between secretory component (SC) and dIgA, through it C-terminal region, which shows particularly high sequence similarity between species (Braathen 2007, Johansen *et al.*, 2001). This therefore gives pIgR/SC its selectivity in only binding polymerised Abs for transportation into the lumen (Brandtzaeg, 1975). These oligomeric Abs are so large and protected by SC that they stand a greater chance of resisting proteolytic degradation in the harsh environment of secretions than monomeric Abs would.

1.4 Secretory IgA

The mucosal surfaces are the largest surface area of the human body in contact with environmental Ags. Consequently, mucosal surfaces are permanently exposed to foreign Ags and are a major entry point for pathogens, and therefore they must be efficiently protected (Lamm, 1997). In response to this threat, one of the first lines of defence of the mucosal immune system is SIgA, where SIgA exerts its function of Ag neutralization and clearance (Nagler-Anderson, 2001). SIgA also acts to restrict the indigenous commensal bacterial populations to the lumen (Macpherson & Harris, 2004).

1.4.1 Formation and secretion of SIgA

IgA is produced in mucosal surfaces locally as polymeric IgA (pIgA), which consists of (usually) dIgA, or higher oligomeric forms of IgA, with the IgA monomers bound covalently via the J chain (Johansen *et al.*, 2001; Figure 1.10). In secretions, the IgA1 and IgA2 subclasses form (usually) homodimeric or homo-oligomeric macromolecules. Secretory IgA (SIgA) is a large macromolecule, with a molecular mass of ~425 kDa.

The human polymeric Ig receptor (pIgR) functions as a receptor for all polymeric forms of IgA, as well as pentameric IgM. The pIgR mediates their transport across the mucosal epithelium into external body fluids (Brandtzaeg & Prydz, 1984; Phalipon & Corthésy, 2003).

pIgR is synthesised as an integral membrane protein in the rough endoplasmic reticulum where it is post-translationally modified by the addition of seven N-linked oligosaccharides (Figure 1.11; Figure 1.12). pIgR then travels to the Golgi apparatus. At the trans-Golgi network, pIgR is sorted into vesicles that deliver it to the basolateral surface of the epithelial cell. At the epithelial cell-surface, pIgR can bind to dIgA that is locally produced by plasma cells, most commonly found in the lamina propria underlying the epithelium (Mostov, 1994). The pIgR and bound dIgA are then transcytosed across the cell (Figure 1.11). Firstly, the pIgR-dIgA complex is endocytosed and delivered to endosomes, where the receptor and its ligand move through a series of endocytotic and transcytotic vesicles. At the apical membrane, a currently uncharacterised enzyme cleaves the extracellular domains of pIgR. This

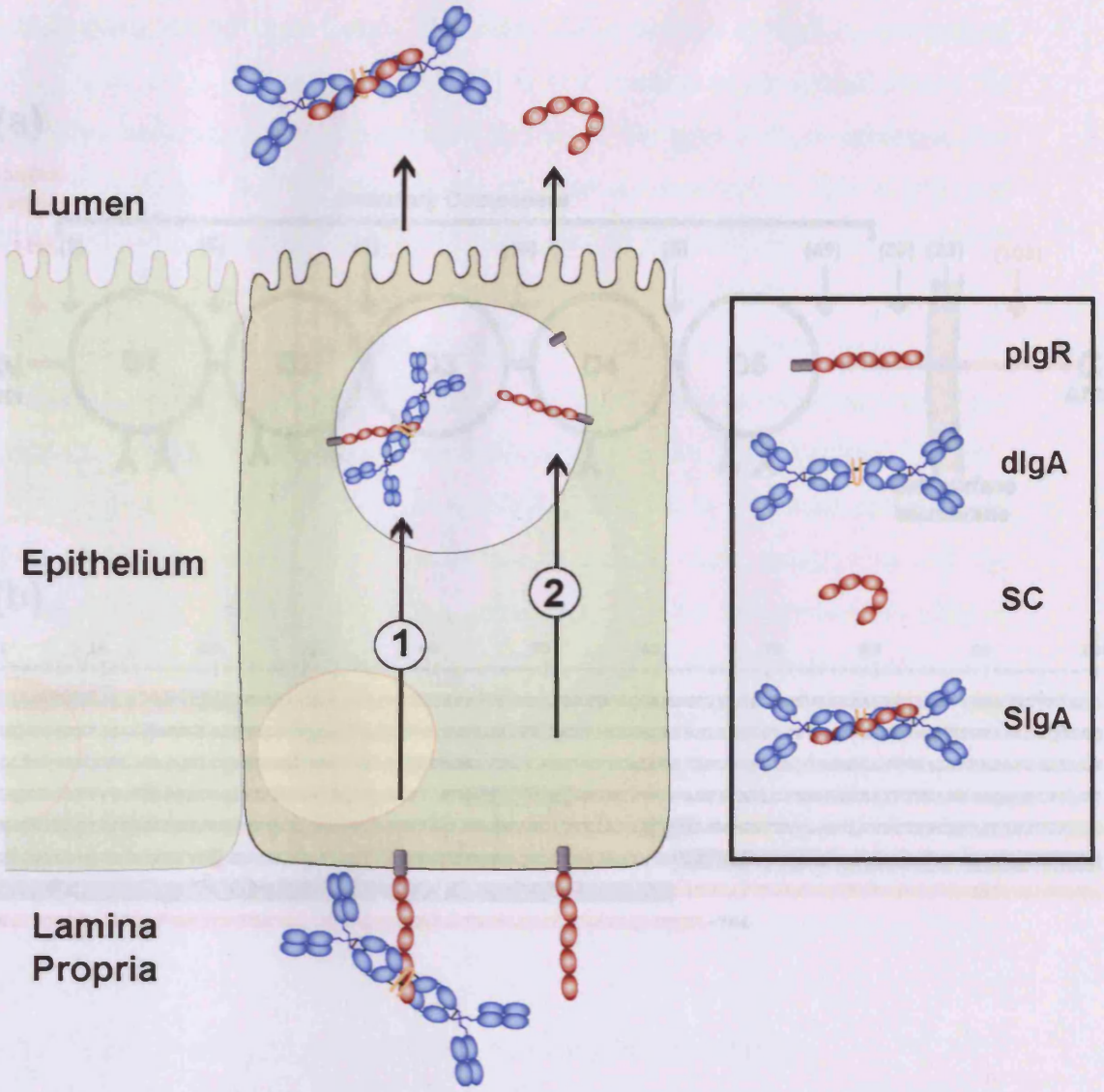

Figure 1.11 Transportation of dIgA across the epithelial lining and formation of SIgA

(1) dIgA in the lamina propria binds to the pIgR. The complex is transcytosed across the cell. The five extracellular domains of pIgR are cleaved and become SC. SC remains bound to dIgA to form SIgA. (2) The pIgR can be transcytosed across the cell without being bound to its ligand dIgA, to release free SC into the lumen.

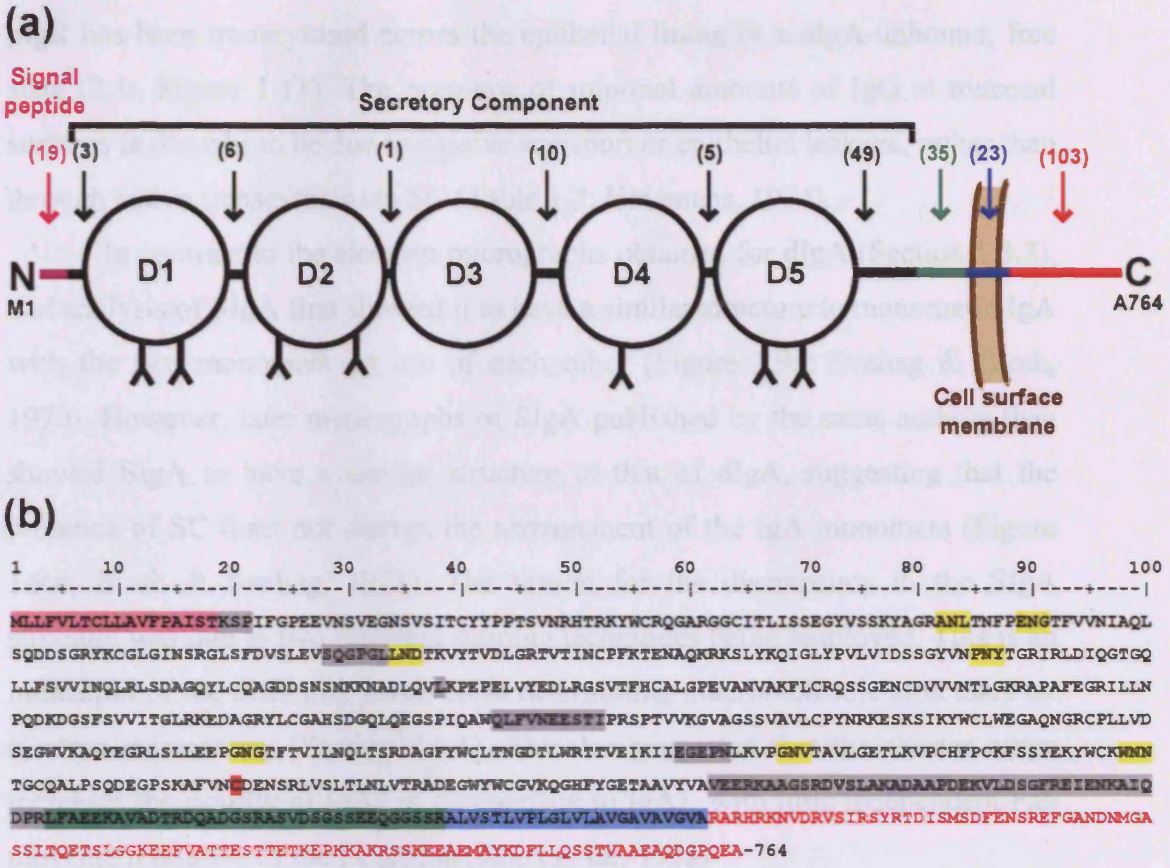

Figure 1.12 Structure and sequence of the polymeric immunoglobulin receptor

(a,b) SC (black) is composed of the first 585 residues of pIgR, after the post-translational removal of the 18-residue signal peptide (pink). SC is formed after the cleavage of the five extracellular domains of pIgR at the apical membrane of mucosal epithelial cells (Figure 1.11). The five extracellular domains of pIgR are labelled D1 to D5. The lengths of inter-domains linkers (bracketed numbers in (a) and grey in (b)), transmembrane region (blue), cytoplasmic tail (red) and seven N-linked glycosylation sites (inverted Y in (a) and yellow in (b)) of pIgR are highlighted. The extracellular portion of pIgR that is not part of SC is shown in green.

extracellular, ligand-binding portion of pIgR remains bound to dIgA, and SIgA is released into the mucosal lumen. The extracellular portion of pIgR is now termed SC (Figure 1.11; Figure 1.12). Free SC is also present in secretions, where the pIgR has been transcytosed across the epithelial lining in a dIgA-unbound, free state (2 in Figure 1.11). The presence of minimal amounts of IgG at mucosal surfaces is thought to be due to passive transport or epithelial leakage, rather than through active transcytosis via SC (Table 1.2; Heremans, 1974).

In contrast to the electron micrographs obtained for dIgA (Section 1.3.3), EM analysis of SIgA first showed it to have a similar structure to monomeric IgA with the two monomers on top of each other (Figure 1.9a; Svehag & Bloth, 1970). However, later micrographs of SIgA published by the same authors then showed SIgA to have a similar structure to that of dIgA, suggesting that the presence of SC does not disrupt the arrangement of the IgA monomers (Figure 1.9.c; Bloth & Svehag, 1971). The reason for the discrepancy in the SIgA structure was due to two different staining techniques being employed. This is an indication of the difficulty involved in determining macromolecule structures by electron microscopy (Section 2.1.1). EM also suggested that the shorter hinge increases the rigidity of IgA2 in comparison to IgA1, with little independent Fab movement relative to the Fc region (Roux *et al.*, 1998).

1.4.2 Secretory component / polymeric immunoglobulin receptor

SC is a glycoprotein consisting of five Ig variable-type domains (D1-D5) to which up to seven carbohydrate chains are attached (Figure 1.12) that account for 15-25% of its total molecular mass (Cleveland *et al.*, 1991). SC comprises the first 585 residues in pIgR (Figure 1.12; Hughes *et al.*, 1997). SC ensures the correct localisation of SIgA in mucosal linings for it to carry out its effector functions. SC bound to dIgA confers protection of the Ab to acid and proteolytic digestion (Crottet & Corthésy, 1998), and governs mucus anchoring of SIgA at mucosal surfaces (Phalipon *et al.*, 2002). The specificity of the pIgR for only polymeric Abs ensures that only those large enough and capable of withstanding the harsh environment of the mucosa are delivered there.

1.4.2.1 Structure of SC domains and their interaction with SIgA

Despite the importance of SC and its glycosylation in many biological and immunological processes, only partial information is available on its structure. Homology modelling of rabbit SC domains 1 and 2 (D1 and D2) indicates that both domains exhibit the typical features of Ig variable (V)-type superfamily members,consisting of seven β-strands, A-G and an extra two designated C' and C'' connected by loops of variable sizes (Williams & Barclay, 1988; Coyne *et al.*, 1994; Corthésy *et al.*, 1996; Section 1.2.2; Figure 1.5). A similar spatial arrangement was determined in the homology modelling of domains D2 and D3 of mouse SC (Crottet *et al.*, 1999). These studies were confirmed by the crystal structure of D1 of human SC, which revealed a V-type fold (Hamburger *et al.*, 2004; Figure 1.13). Three CDR are evident at one end of SC D1 (Figure 1.13).

SC binds to dIgA with a 1:1 stoichiometry (Crotett & Corthésy, 1999). SC D1 contains the three CDR, which are thought to contribute to the initial binding of SC to dIgA (Frutiger *et al* 1986; Bakos *et al.*, 1991b). D2 and D3 are known to be necessary in SIgA formation, yet do not establish close contact with the dIgA backbone (Crotett & Corthésy, 1999). D5 contains Cys502, which is the covalent binding site to Cys311 in one monomer in dIgA (Figure 1.10; Figure 1.12; Fallgreen-Gebauer *et al.*, 1993). It is thought that the initial contact of D1 with the dimer triggers a 'zipper effect', where contact between SC and dIgA moves from D1 down the domains to bring D5 into correct alignment for disulphide bond formation (Bakos *et al.*, 1991b; Crotett & Corthésy, 1999). The binding between SC and dIgA also involves the J chain (Johanson *et al.*, 2001). The glycosylation of SC is not thought to be relevant to its binding to dimeric IgA (Bakos *et al.*, 1991b). Instead, the carbohydrates on SC give it its own immunogenic role and protection to itself and SIgA (Perrier *et al.*, 2006). The binding of SC to the different SIgA isotypes may differ, with SC covalently binding to SIgA1 yet interacting with SIgA2 in a non-covalent manner only (Almogren *et al.*, 2007).

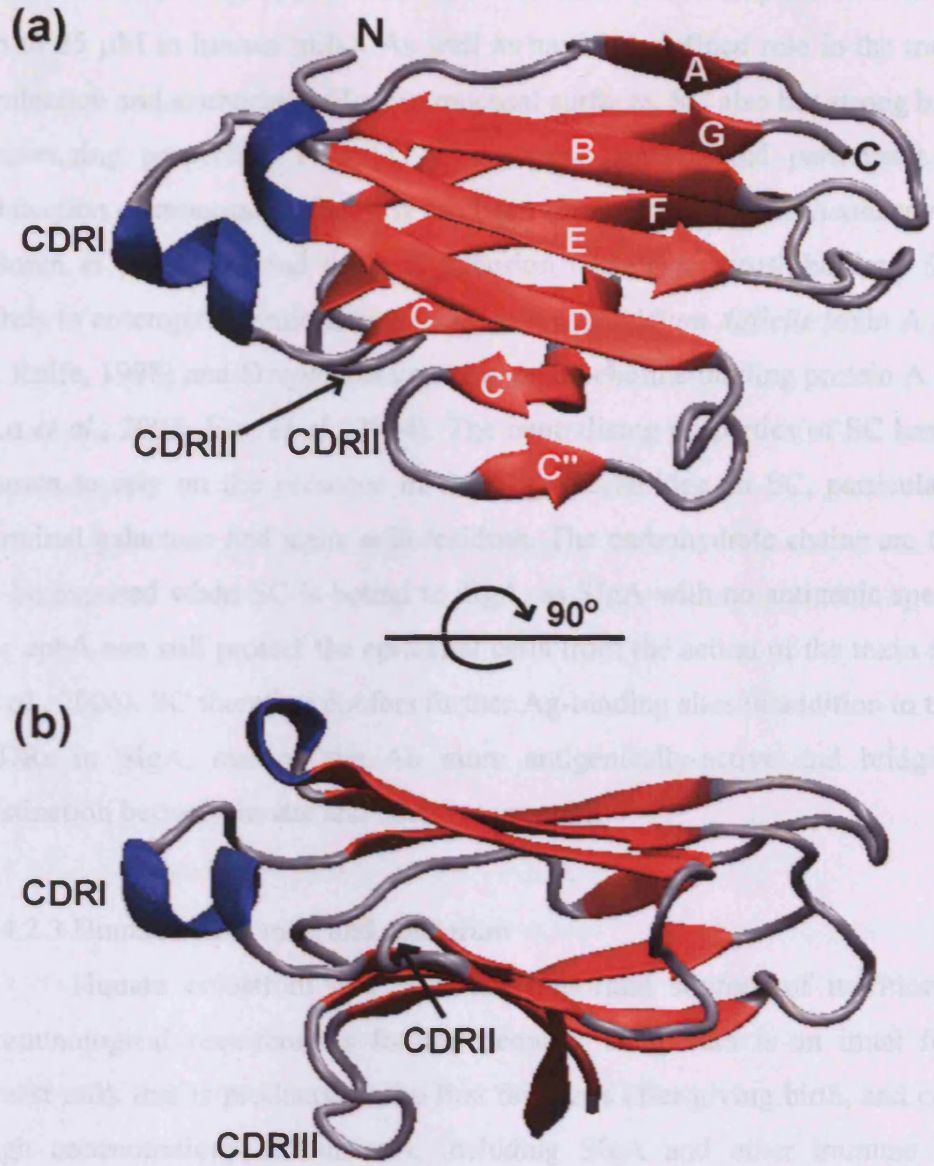

Figure 1.13 Crystal structure of secretory component domain one

The crystal structure of SC domain one (D1) at 0.19 nm is shown (PDB 1XED; Hamburger *et al.*, 2004). The β-strands (red), alpha helices (blue), loop regions (white), N- and C-terminals and complementarity determining regions (CDRs) are highlighted. The β-strands are labelled in (a) and the D1 is orientated to show the DEBA face to the front. In (b) D1 has been rotated 90° on the X-axis in respect to (a).

1.4.2.2 Functions of free SC

SC is present in secretions in a free form (for example at concentrations up to 25 μM in human milk). As well as having a defined role in the transport, protection and anchoring of IgA in mucosal surfaces, SC also has strong bacterial scavenging properties. Free SC is an active antibacterial participant in the protection of mucosal surfaces. It has been shown to bind to *Helicobacter pylori* (Boren *et al.*, 1993) and limit its adhesion to hamster brush-borders. SC also binds to enteropathogenic *Escherichia coli*, *Clostridium difficile* toxin A (Dallas & Rolfe, 1998) and *Streptococcus pneumoniae* choline binding protein A (cbpA) (Lu *et al.*, 2003; Elm *et al.*, 2004). The neutralising properties of SC have been shown to rely on the presence of the oligosaccharides on SC, particularly the terminal galactose and sialic acid residues. The carbohydrate chains are thought to be exposed when SC is bound to dIgA, as SIgA with no antigenic specificity for cpbA can still protect the epithelial cells from the action of the toxin (Perrier *et al.*, 2006). SC therefore confers further Ag-binding sites in addition to the four CDRs in SIgA, making the Ab more antigenically-active and bridging the distinction between innate and adaptive immunity.

1.4.2.3 Human breast milk and colostrum

Human colostrum and milk are important sources of nutritional and immunological requirements for the neonate. Colostrum is an inital form of breast milk that is produced in the first few days after giving birth, and contains high concentrations of nutrients, including SIgA and other immune system components to help protect the neonate and the lactating mammary glands from pathogenic microorganisms. Unlike other external secretions, human milk contains viable leukocytes found at highest concentrations in colostrum (Brandtzaeg, 1983). Human milk contains a complex mixture of aqueous and cellular components. The aqueous components include SIgA, free SC, lactoferrin, casein, α-lactalbumin and lactose. The presence of SIgA and free SC in colostrum and milk are an example of the combination of adaptive and innate immune system working together for maximum effectiveness. The mother therefore provides the neonate with adaptive immune system components and

therefore Ag-specificity through SIgA, as well as though the delivery of IgG across the placenta to the foetus (Figure 1.3).

1.4.3 IgM as an accessory secretory immunoglobulin

Pentameric IgM can be present in external secretions with an accessory role as a secretory immunoglobulin. This suggests that SC has a generic Ab-binding ability, being able to bind to both dIgA and polymeric IgM (Mach, 1970). The J chain may confer this generic polymeric Ab-binding ability to SC, as the J chain is known to contribute to the SC binding site on polymeric immunoglobulins, and may give SC its selectivity in only binding those immunoglobulins that are polymerised and ready for transportation into the lumen (Brandtzaeg, 1975). This selectivity may be required to only permit larger, polymerised Abs, with protection in the form of heavily glycosylated SC, to enter the harsh environment of external secretions. Smaller, monomeric immunoglobulins would be readily subjected to proteolysis and would consequently be unable to fulfil their role of pathogen exclusion.

1.5 Mucosal immunity and the SIgA response

The mucosal surfaces in the human body, including the gastrointestinal, respiratory and urogenital tracts represent around 400 m^2. Mucosal surfaces are therefore the largest surface area in the human body in contact with the external environment (compared with 1.8 m^2 for skin) (Childers *et al.*, 1989). Pathogenic microorganisms and their potentially toxic products and antigens present in ingested food and inhaled air provide a constant threat to mucosal surfaces.

Mucosal surfaces possess powerful specific and non-specific defence mechanisms to prevent the entry of pathogens and exogenous materials invading the human body and to maintain local homeostasis. Mucosal surfaces have physical and mechanical barriers, such as a wall of epithelial cells, mucus and peristalsis. Also, the action of Abs, mucosal-associated lymphoid tissue (MALT), immunoreactive cells present in mucosal tissue and the populations of commensal microbiota all contribute to mucosal immunity (Corthésy, 2007). SIgA is the principle Ab class produced in the mucosal tissues (Mestecky & McGhee, 1987), with a total daily synthesis of IgA around 66 mg/kg representing

more than 60% of the total immunoglobulin produced (Childers *et al.*, 1989). Fine-tuning of all these effector mechanisms regulate the immune responses, oral tolerance and immune ignorance, to optimise the specific immune response and prevent local tissue damage.

1.5.1 Mucosal Associated Lymphreticular Tissue (MALT)

The mucosal immune system is considered distinct from the systemic humoral immune system that protects the rest of the human body. This concept was first proposed after the analysis of the immune system of rodents in 1919. It was shown that rabbits, which had been orally immunised with denatured *Shigellae*, were protected against fatal dysentery regardless of serum Abs. In the 1930s, this concept was supported after the identification that the upper respiratory tract was also part of the mucosal system (Mestecky *et al.* 1999). This has since been confirmed in humans.

The mucosal immune system is composed of mucosal-associated lymphoid tissue (MALT), which includes the gut-associated lymphoid tissues (GALT) and the bronchial-associated lymphoid tissues (BALT) (Mestecky & McGhee, 1987). Collectively, MALT is organised lymphoid tissue that contains lymphocytes necessary to protect the mucosal immune system from all Ags.

GALT is one of the most intensively studied mucosal membranes. GALT is a highly evolved system that not only protects the host from ingested pathogens, but it has also developed mechanisms to prevent the host from reacting against ingested proteins and itself (Weiner *et al.*, 1994). Organisation of some of the components of GALT is shown in Figure 1.14. A specialised region of cells, called Peyer's patches (PP) are located in the submucosal layer of the gastrointestinal tract. In humans, they are described as macroscopic lymphoid aggregates that can be seen by the naked eye on the serosal side of the small intestine, with the greatest intensity in the ileum. PP contain clusters of lymphocytes, and are surrounded by microfold (M) cells. M cells differ from normal epithelial cells in that they are flatter, lack microvilli and have large invaginations of the basolateral membrane to form pockets. Lymphocytes are present in these M cell pockets, including B cells, T cells and APCs such as dendritic cells (DC) and macrophages.

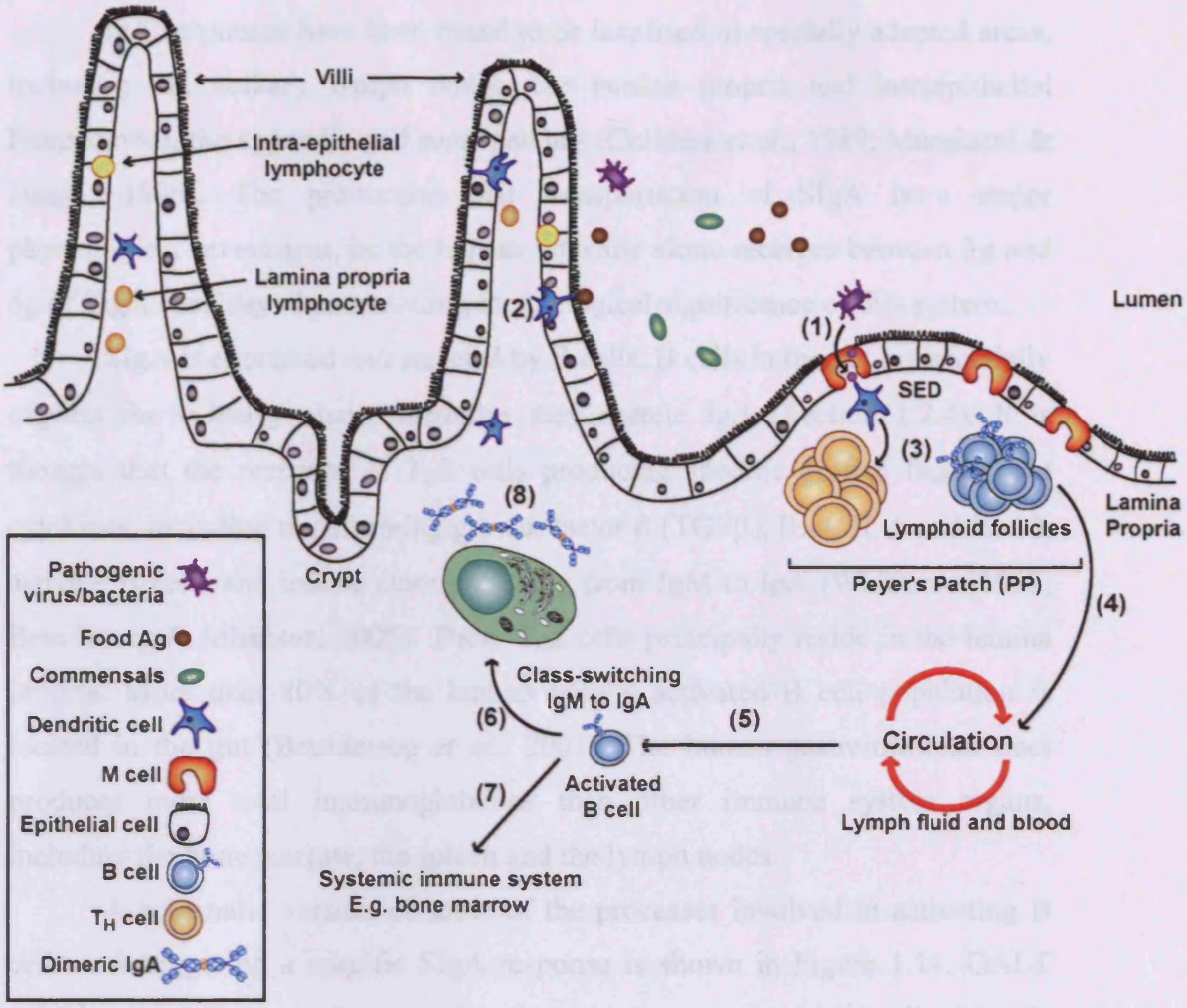

Figure 1.14 Antigen processing and mucosal immune response induction

(1) Pathogens enter a Peyer's patch (PP) via M cells, where they are processed by immature dendritic cells (DC) in the sub-epithelial dome (SED) region. (2) Intraepithelial DC can also directly contact and process Ag. (3) T cells in lymphoid follicles are then activated. (4) B cells are activated by cytokines released by the T cells and (5) differentiate to mature B cells whilst circulating the body in the lymph fluid and the blood stream. (6) Most of the activated B cells return to effector sites, while (7) some migrate to systemic immune system organs. (6) In GALT, activated B cells preferentially express IgA by class-switching their heavy chain Ig gene from IgM. (8) Plasma cells in the lamina propria produce pIgA to be transported across the epithelial cells for release into the lumen as SIgA to bind to and remove the Ag.

1.5.2 The role of SIgA in mucosal defence

IgA responses have been found to be localised in specially adapted areas, including PP, solitary lymph nodes, the lamina propria and intraepithelial lymphocytes, the appendix and cecal patches (Childers *et al.*, 1989; Murakami & Honjo, 1995). The production and transportation of SIgA is a major physiological investment, as the human intestine alone receives between 3g and 5g of SIgA each day, demonstrating the biological significance of this system.

SIgA is expressed and secreted by B cells. B cells in the gut preferentially express the α-heavy chain, therefore they secrete IgA (Section 1.2.4). It is thought that the presence of T_H2 cells producing specific growth factors and cytokines, including transforming growth factor β (TGFβ), IL-4, IL-5 and IL-10, activate B cells and induce class switching from IgM to IgA (Whitmore, 1992; Brandtzaeg & Johansen, 2005). These T_H2 cells principally reside in the lamina propria. More than 80% of the human body's activated B cell population is located in the gut (Brandtzaeg *et al.*, 2001). The human gastrointestinal tract produces more total immunoglobulins than other immune system organs, including the bone marrow, the spleen and the lymph nodes.

A schematic version of some of the processes involved in activating B cells and triggering a specific SIgA response is shown in Figure 1.14. GALT exhibits the ability to endocytose Ags from the lumen, through M cells. M cells can phagocytose a luminal pathogen and then active APCs, such as DC, to generate an immune response against that Ag. DC are present in the sub-epithelial dome (SED) to sample Ag from the lumen, and control the activation of naive T cells present in lymphoid follicles in PP. DC can also directly carry out Ag processing and presentation, by extending dendrites between epithelial cells to directly contact and capture Ag in the lumen. Activated B and T cells then emigrate from the inductive environment via the lymphatic system, and circulate in the lymph fluid and the bloodstream. Activated B cells travel to effector sites, represented by the lamina propria. During this migration, B cells differentiate to IgA-producing B cells, through class switching of the Ab heavy chain (Section 1.2.4). SIgA can then be secreted with the capability of specifically recognising and eliminating the Ag, after its trancytosis across the epithelial cell into the lumen (Section 1.4.1; Figure 1.11; Corthésy, 2007).

Once in the mucosal lumen, SIgA can bind Ag and then can adhere to the apical surface of M cells. The IgA-receptor complex is then endocytosed, as seen by gold-coated luminal SIgA molecules being present in the pockets of M cells (Weltzin *et al.*, 1989). The endocytosis of SIgA by M cells is one of the main routes to deliver pathogens to mucosal immune system inductive sites (Figure 1.14). SIgA is thought to bind to M cells by interacting with a currently undefined receptor on the M cell surface, which is thought to bind the SIgA C_H1 and C_H2 domains and/or the hinge region (Mantis *et al.*, 2002).

SIgA also prevents pathogens invading the mucosa by 'immune exclusion', where SIgA prevents the attachment and penetration of microorganisms rather than their ultimate destruction. The four or more Ag-binding sites on SIgA block any adherence-determinants on the pathogen and facilitate their removal through muco-cilary flow or peristalsis. The presence of SC and the carbohydrate moieties are thought to be important in their anchorage to mucosal surfaces (Phalipon *et al.*, 2003; Perrier *et al.*, 2006).

1.5.3 Mucosal microflora and SIgA in mucosal tolerance

The tracts of the human body, particularly the proximal and distal portions, are continuously colonised by commensal organisms. While the presence of commensal bacteria generally prevents the colonisation of pathogens through competitive exclusion, they represent an enormous antigenic-challenge to the mucosal immune system. Mucosal or oral tolerance is the down-regulation, hypo-responsiveness or suppression of immune effector cell responses to an Ag after prior administration of the Ag by the same route. Mucosal tolerance is also important in the discrimination of self from non-self, as if the immune system reacted against self, an autoimmune disease would result. IgA and other immune responses to commensal bacteria are down-regulated by mechanisms such as clonal anergy of B and T cell populations and the secretion of suppressive cytokines (Weiner, 1994). It is thought that IgA may prevent immune responses by excluding mucosal Ags from tissues. Mice with no pIgR (therefore significantly decreased levels of mucosal IgA) showed signs of mucosal inflammation from unnecessary triggering of systemic immunity against innocuous Ag and indigenous bacterial flora, as mucosal tolerance could not be

initiated (Johansen *et al.*, 1999). The lack of mucosal tolerance in thought to be a significant immunopathogenic mechanism in inflammatory diseases, such as celiac disease and irritable bowel syndromes. The GALT therefore cannot be used as the site of immunisation for vaccinations, as mucosal tolerance can be naturally generated (Weiner, 1994).

1.6 IgA receptors

The IgA receptor family consists of a number of cell-surface receptors, including the pIgR (Section 1.4.2), the IgA Fc receptor (FcαRI or CD89) (Figure 1.15) and the Fcα/μR. At least two other IgA receptors have been identified, including the asialoglycoprotein receptor and the transferrin receptor.

Fc receptors (FcR) belong to the immunoreceptor family, which includes T-cell, B cell and NK receptors. FcR are present on many immune system cell types and they provide a crucial link between the humoral and cell-mediated branches of the immune system. FcR are transmembrane receptors that recognise the Fc region of immunoglobulins. FcR for all five-antibody classes have been described. The interaction between Abs and their FcR provides Ag-specific recognition to cells that express the specific FcR. Therefore, through this Fc-FcR interaction a variety of responses can be initiated, including endocytosis, exocytosis, transcytosis, respiratory bursts, degranulation, antibody-dependent cell cytotoxicity (ADCC), cytokine release, superoxide generation and phagocytosis (Monteiro & van de Winkle, 2003).

1.6.1 IgA Fcα receptor

IgA induces immune effector functions by binding to and activating the IgA-specific receptor IgA Fcα receptor I (FcαRI, also called CD89). FcαRI is present on many immune cell types, including monocytes, neutrophils, macrophages, intestinal dendritic cells and eosinophils (Kerr & Woof, 1999). It is estimated that FcαRI-expressing cells contain between 6000 and 7000 molecules of the receptor per cell (Morton *et al.*, 1996).

The FcαRI is a transmembrane glycoprotein. In its mature form (after the removal of the 21-residue signal peptide), FcαRI consists of 266 amino acids,

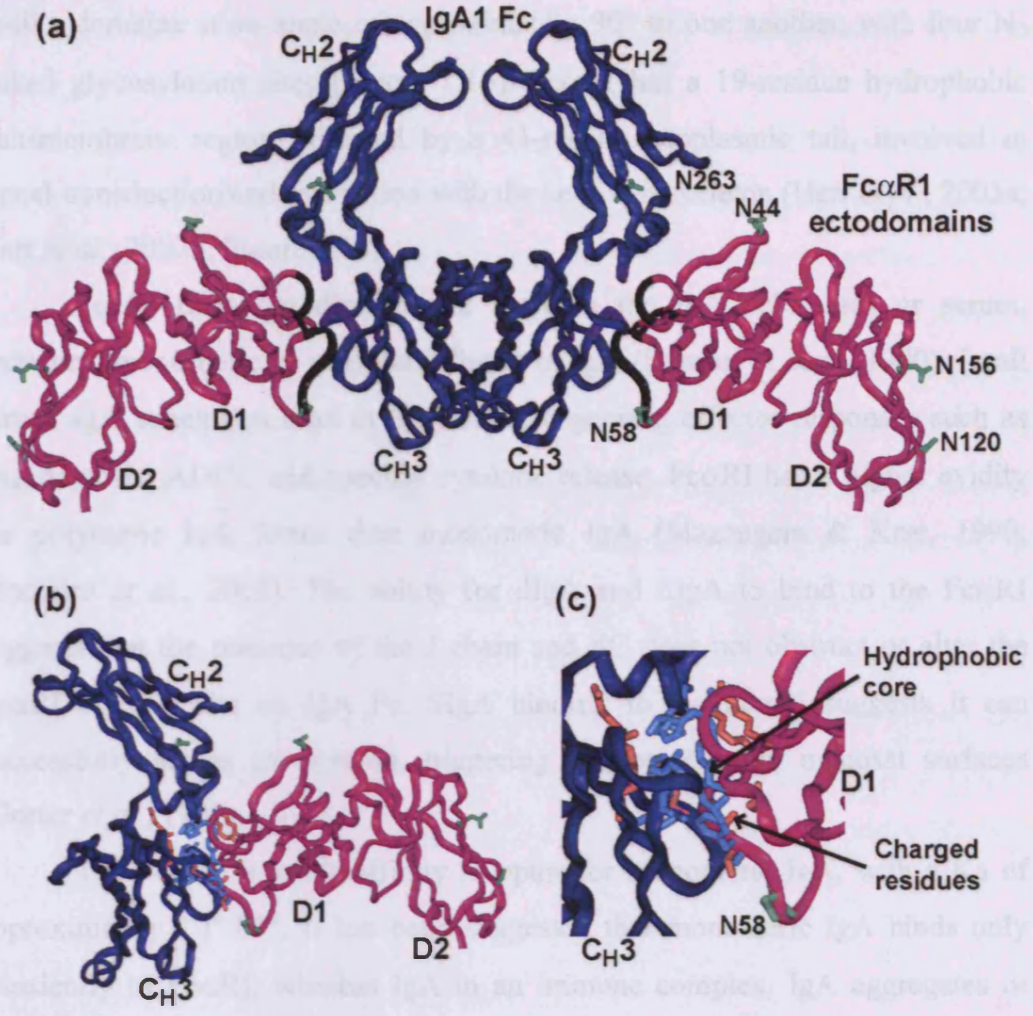

Figure 1.15 Interaction between the Fc regions of IgA and FcαRI

The D1 and D2 ectodomains of the FcαRI (pink), C_H2 and C_H3 IgA1 Fc domains (blue), asparagine residues constituting N-linked carbohydrate sites (green) are shown (PDB 1OW0; Herr *et al.*, 2003b). (a) The regions involved in Fc-FcαRI interactions are black (FcαRI) and purple (IgA Fc). (b,c) The residues involved in Fc-FcαRI interactions are shown (sticks), where light blue residues constitute the hydrophobic core, and the charged residues flanking are shown in orange.

with a predicted protein molecular mass of around 30 kDa. However, due to differential glycosylation, FcαRI has a molecular mass between 55 and 75 kDa (Morton *et al.*, 1996). The 206-residue FcαRI extracellular portion contains two Ig-like domains at an angle of approximately 90° to one another, with four N-linked glycosylation sites (Figure 1.15). FcαRI has a 19-residue hydrophobic transmembrane region followed by a 41-reside cytoplasmic tail, involved in signal transduction and interaction with the cells cytoskeleton (Herr *et al.*, 2003a; Herr *et al.*, 2003b; Figure 1.15).

FcαRI does not discriminate between the IgA subclasses or serum, secretory, monomeric or polymeric forms of IgA (Stewart & Kerr, 1990). In all forms, IgA subclasses bind to the FcαRI, triggering effector responses such as phagocytosis, ADCC and specific cytokine release. FcαRI has a higher avidity for polymeric IgA forms than monomeric IgA (Mazengera & Kerr, 1990; Monteiro *et al.*, 2002). The ability for dIgA and SIgA to bind to the FcαRI suggests that the presence of the J chain and SC does not obstruct or alter the FcαRI binding site on IgA Fc. SIgA binding to the FcαRI suggests it can successfully act as an opsonin, triggering phagocytosis on mucosal surfaces (Gorter *et al.*, 1987).

The FcαRI is a low-affinity receptor for monomeric IgA, with a Ka of approximately 10^6 M^{-1}. It has been suggested that monomeric IgA binds only transiently to FcαRI, whereas IgA in an immune complex, IgA aggregates or higher polymeric IgA forms bind more efficiently (Wines *et al.*, 1999; Stewart & Kerr, 1990). Also, monomeric IgA dissociates more rapidly from the receptor than dIgA, altering the effects and signalling potency from the two IgA forms (Oortwijn, *et al.*, 2007).

The stoichiometry of FcαRI binding to monomeric IgA is a 2:1 complex, where 2 receptors bind to each C_H2-C_H3 heavy chain in the Fc region (Figure 1.15; Herr *et al.*, 2003a). The crystal structure of human FcαRI ectodomain alone and in complex with an IgA1 Fc region has been determined at 0.30 nm and 0.31 nm respectively (PDB code 1OVZ and 1OW0; Herr *et al.*, 2003b). The binding interaction includes hydrophobic resides forming a hydrophobic core, which is flanked by charged residues, in IgA1 Fc and the FcαRI (Figure 1.15). It has been shown that IgA carbohydrates do not participate in the interaction with its

FcαRI, as the absence of the IgA N-linked carbohydrates in the C_H2 constructs does not impede it binding to FcαRI on neutrophils (Mattu *et al.*, 1998; Herr *et al.*, 2003b). However, the carbohydrate moiety attached to FcαRI at Asp58 is involved in the IgAFc- FcαRI interaction, forming two hydrogen bonds and van der waals contacts with the IgA Fc (Herr *et al.*, 2003b).

FcαRI is expressed either associated, or non-associated with a signalling receptor, called FcR γ-chain, which contains an immunoreceptor tyrosine-based activation motif (ITAM). Little is known about the residues involved in the interaction of FcαRI with FcR γ-chain, yet they are known to interact strongly and signal via tyrosine kinases Lyn, Syk and Btk (Gulle *et al.*, 1998; Launay *et al.*, 1998). When present together, the FcαRI signalling through ITAM is a mechanism by which FcαRI can carry out cellular functions such as calcium release, Il-2 secretion, degranulation or IgA degradation (Monterio & van de Winkel, 2003). FcαRI has been implicated in the pathogenesis of IgAN (Monteiro *et al.*, 2002). A soluble form of the FcαRI with a molecular mass of around 30 kDa has been identified in human serum. The soluble FcαRI can covalently link to pIgA (van de Boog *et al.*, 2002). This soluble form of FcαRI may therefore contribute to IgA aggregation in IgAN mesangial deposits (Section 1.7.2).

1.6.2 IgA Fcα/μ receptor

A 503-residue mouse Fcα/μ receptor (Fcα/μR) has been described, which contains one extracellular domain and four potential N-linked glycosylation sites. A human homologue has also been identified from sequence similarities, with the ability to bind both IgA and IgM (Monteiro & van de Winkle, 2003). The Fcα/μ receptor on B-cells and macrophages internalises IgM and IgM-coated microbes, and is thought to play a role in the primary stagess of an antimicobial immune responses after B cell stimulation (Shibuya *et al.*, 2000). Also, Fcα/μR mRNA transcripts have been shown to be present in human mesangial cells, suggesting that this receptor may be implicated in the pathology of IgAN (McDonald *et al.*, 2002; Section 1.7.2).

1.6.3 Secretory component receptor

Secretory component receptor (SCR) binds SC and SIgA but not serum IgA. SCR is expressed on eosinophils and basophils and has a molecular mass of 15 kDa. The binding of SC or SIgA to SCR is thought to trigger degranulation and the release of eosinophil cationic protein and peroxidase. Therefore, as well as FcαRI, SCR is another receptor involved in mucosal immunity specific for SIgA, which may facilitate the potency of SIgA to trigger eosinophil degranulation (Lamkhioued *et al.*, 1995).

1.6.4 Transferrin receptor

Transferrin receptor (TfR or CD71) is involved in cellular iron metabolism. TfR is expressed on a wide range of tissues, including renal mesangial cells. A second isoform of this receptor have been identified (TfR2), which has only been identified on liver and erythroid cells (Vogt *et al.*, 2003; Decot *et al.*, 2005). TfR has been identified unexpectedly as an IgA receptor, where it selectively binds IgA1 and preferentially monomeric IgA1 over polymeric IgA1 or SIgA1. This selectivity of TfR for IgA1 may be through the unique O-linked carbohydrates on the IgA1 hinge region (Moura *et al.*, 2001). The O-linked carbohydrate moieties on the IgA1 hinge are thought to be altered on the IgA molecules in IgAN deposits (Section 1.7.2). Nevertheless, TfR is thought to be one of the receptors responsible for binding IgA1, leading to the IgA-depositions in the kidney in IgAN (Moura *et al.*, 2005).

1.6.5 Asialoglycoprotein receptor

The liver is known to play an important role in maintaining IgA homeostasis, principally through the regulation of IgA catabolism. The asialoglycoprotein receptor (ASGP-R) is expressed on hepatocytes. ASGP-R recognises terminal galactose residues in carbohydrate moieties on serum glycoproteins, including IgA. Liver IgA catabolism may involve ASGP-R, as ASGP-R is proposed to be involved in IgA clearance from circulation (Stockert, 1995). The major clearance pathway for the IgA2 isotype is mediated by hepatocytes ASGP-R, whereas only a small percentage of IgA1 is cleared via this pathway and receptor (Rifai *et al.*, 2000). This may explain the higher proportion

of IgA1 in blood serum relative to IgA2, and may be involved in elevating serum IgA1 levels in IgAN.

1.7 Immunoglobulin A in disease

1.7.1 Selective IgA deficiency

More IgA is produced than all the other antibody classes combined, yet remarkably, many individuals with IgA deficiency, the most prevalent immunodeficiency of all, remain relatively healthy. The prevalence of IgA deficiency ranges from 1:223 to 1:1000 (Hanson, 1983), and is characterised as undetectable serum IgA levels in the presence of normal serum levels of IgG and IgM. IgA deficiency may be familial or acquired, following the parasitic disease toxoplasmosis, measles, or other viral infections. Patients with IgA deficiency are generally asymptomatic, but show increased susceptibility to infections, including sinopulmonary tract infections such as pneumonia and sinusitis, and gastrointestinal infections, particularly celiac disease and inflammatory bowel disease (Cunningham-Rundles, 2001). Patients with IgA deficiency also have an increased chance of developing autoimmune diseases in middle age. The lack of disease and symptoms in IgA deficiency is attributed to a compensatory increase in secretory IgM. However, although IgM is a functionally active secretory immunoglobulin, it is thought to be unable to confer the same mucosal protection as IgA (Savilahti *et al.*, 1988). In the absence of IgA, a significant proportion of patients with IgA deficiency have anti-IgA antibodies, which can lead to severe reactions including anaphylaxis if a blood-transfusion is attempted (Cunningham-Rundles, 2001).

1.7.2 IgA nephropathy

IgA nephropathy (IgAN, also known as Berger's disease) is the most common form of chronic glomerulonephritis (an inflammation of the glomeruli of the kidney) in the world. IgAN was first described in 1968 (Berger and Hinglais, 1968). Primary IgAN is characterised by mesangial depositions of IgA, particularly polymeric and dimeric IgA1 and increased serum levels of dIgA1 (Barratt & Feehally, 2005). Since dIgA1 is predominantly mucosally-derived, it was thought that the IgA deposits may arise from over-production of mucosal

IgA abnormally entering circulation. However, the deposited IgA is now known to originate in the bone marrow (Feehally & Allen, 1999).

IgAN symptoms include renal impairment, haematauria, proteinuria, upper respiratory tract infection and hypertension. IgAN can frequently progress to acute or chronic renal failure (Barratt *et al.*, 2007). Hypercomplementemia has been seen in IgAN patients, with mesangial co-deposits of the complement components C3 and properdin (Onda *et al.*, 2007). Also, IgA is present in association with MBL and C4d (Oortwijn *et al.*, 2007). A variety of systemic disorders are associated with (secondary) IgAN, including liver disease, coeliac disease, rheumatoid arthritis and HIV infection and AIDS. Reoccurrence of IgAN in patients after transplantation of a normal kidney suggests that IgAN is caused by an abnormality related to the structure of IgA, IgA production and/or an IgA receptor rather than an intrinsic abnormality of the kidney.

Alterations in the glycosylation patterns of IgA may be involved in the formation of IgA aggregation in mesangial deposits. A reduced galactose (Gal) and sialic acid (Neu5Ac) content, increased exposure of N-acetylgalactosamine (GalNAc) and increased O-glycosylation has been identified on circulating IgA1 in IgAN (Coppo & Amore, 2004). Exactly how the IgA molecules are incorrectly glycosylated is currently poorly understood. IgA with truncated O-linked carbohydrates may arise from a specific B cell lineage secreting the incorrectly glycosylated IgA into circulation (or correctly glycosylated IgA into the wrong area of the human body), or abnormalities may occur in the control of IgA glycosylation (Barratt *et al.*, 2007).

Deposits may also form through IgA binding to an over-expressed and/or abnormal receptor(s) present on the mesangial cells surface, where FcαRI, Fcα/μR and/or the TfR are possible candidates (Section 1.6.1; Section 1.6.2; Section 1.6.4). Human mesangial cells can bind IgA yet, until recently, were not known to express an IgA receptor. Both the Fcα/μR and the TfR receptor have been shown to be over-expressed in IgAN patients (McDonald *et al.*, 2002; Moura *et al.*, 2001). mRNA encoding the Fcα/μR was identified in mesangial cells. The receptor was found to bind to IgA and IgM, but not IgG (McDonald *et al.*, 2002), suggesting the receptor-Ab interaction involves a generic binding site on IgA and IgM that is absent in IgG, possibly the tailpiece. However, IgAN

deposits primarily contain dIgA1 or polymeric IgA1, of which the binding ability to Fcα/μR was not tested. The over-expression of the TfR and its co-localisation with IgA1 in mesangial cells was found to be induced by the serum of IgAN patients, but not healthy individuals (Moura *et al.*, 2005).

Soluble FcαRI may be involved in the IgA aggregation leading to the IgA deposition in IgN. Soluble FcαRI in complex with IgA has identified in the serum of IgAN patients (Launay *et al.*, 1999; Monteiro *et al.*, 2002). However, such complexes have not been identified in patients with other IgA-related diseases. Also, soluble FcαRI has been identified in the serum healthy individuals with no IgA depositions characteristic in IgAN (van de Boog *et al.*, 2002). This suggests a combination of altered IgA structure, glycosylation and/or production as well as altered IgA FcαRI may be required for IgAN development.

1.7.3 Other diseases associated with IgA

Elevated levels of serum IgA are commonly associated with a number of specific diseases. The liver is thought to play an important role in the metabolism of IgA, through the asialoglycoprotein present on hepatocytes (Section 1.6.5), yet its exact function is unclear. Serum IgA concentrations are markedly increased in liver disease, including polymeric and secretory IgA. The cause of this increase in IgA levels is not known. Raised levels of IgA, IgA-containing immune complexes and/or IgA rheumatoid factor are found in individuals suffering from alcoholic cirrhosis, HIV infection and AIDS and rheumatoid arthritis (Deviere *et al.* 1992; Grossetette *et al.* 1995; Otten *et al.* 1992; Woof & Kerr, 2006).

IgA-deficient individuals have an increased incidence of autoimmune diseases, particularly associated with the gastrointestinal tract. Coeliac disease is a genetically predisposed, autoimmune disease afflicting the small intestine in individuals of all ages. Coeliac disease is caused by an immune reaction against gliadin, a wheat glycoprotein. Upon exposure, gliadin is modified by transglutaminase and consequently an immune inflammatory reaction occurs (Robins & Howdle, 2005). Patients with selective IgA deficiency have at least a tenfold risk of coeliac disease (Collin *et al.*, 1992). Coeliac disease is often associated with dermatitis herpetiformis, an IgA-mediated chronic, extremely

itchy rash consisting of papules, which are solid, conical elevations of the skin (Fry, 2002).

1.8 Therapeutic uses of Immunoglobulin A

Dysregulation of antibodies can lead to formation of autoantibodies and subsequently autoimmune diseases. In some cases intravenous immunoglobulins can be administered as replacement therapy in patients with immunodeficiencies. However, anaphalaxis can occur, most commonly in IgA deficiency (Pyne *et al.*, 2002).

IgA antibodies are a useful tool in the diagnosis of certain diseases affecting the gastrointestinal tract. For example, elevated levels of serum IgA anti-*Saccharomyces cerevisiae* antibodies have been detected in patients with Crohn's disease. An increase in these IgA antibodies is specific for Crohn's disease, therefore their detection by screening may help towards diagnosis, as increased IgA production is seldom seen in other inflammatory bowel diseases (Barnes *et al.*, 1990).

The pIgR is a potential vehicle for biologically-active molecules or drugs to specifically deliver them to epithelial cells or across epithelial linings to mucosal surfaces. For example, a binding site thought to be on domains three and/or four of the pIgR, which specifically binds to an 11-residue peptide has been identified. The peptide binding site must be distinct from the region on pIgR that interacts with its ligands, as the peptide must not compete with dIgA or pentameric IgM (Braathen *et al.*, 2006). Similarly, a therapeutic strategy has been considered for the treatment of cyctic fibrosis, exploiting the pIgR-mediated cellular route. Agents for delivery have been linked to mouse monoclonal Abs, raised against the extracellular domains of pIgR for delivery across epithelia (Ferkol *et al.*, 1995). The pIgR therefore may be a promising receptor for the delivery of specific drugs or gene therapy agents, to treat diseases that specifically affect mucosal surfaces.

SIgA-based Abs against *Steptococcus mutans* adhesion proteins have been shown to offer host protection against oral streptococcal-recolonisation, showing that IgA Abs may be beneficial in caries control and other oral or mucosal infections (Ma *et al.*, 1998).

In addition to surgery, chemotherapy and radiotherapy, antibody-based immunotherapy is a novel treatment option for some oncological patients. This could be through direct or indirect actions. Direct effects are mediated by the Ab after binding to the target in order to inhibit growth, induce apoptosis or arrest the cell cycle and indirect effects are mediated by the host's immune system and require the Ab to interact with cellular Fc receptors or complement components (Dechant & Valerius, 2001). The role of IgA as an antibody for immunotherapy is not well investigated, but it is thought to have many advantages over IgG, the currently most widely used antibody isotype for cancer therapy. For example, IgA in its polymeric form has an increased valency, therefore potentially increasing signalling capacities. Also, IgA could be utilised to attack common tumours of the lung and gastrointestinal tract due to its specific delivery to the mucosal surfaces (Sato *et al.*, 1997).

Chapter Two

Methods to Determine Antibody Structures

2.1 Introduction to antibody structure determination methods

In order to understand the functions of a protein at a molecular level, it is often necessary to determine its three-dimensional structure. There are many protein structure determination techniques that give different yet often complementary information about the structure of proteins. Protein amino acid primary structures can be resolved directly by mass spectrometry or Edman degradation, or indirectly through the determination of the nucleotide sequence of the protein gene or cDNA. Circular dichroism and Fourier transform infrared spectroscopy are two low-resolution techniques that can be employed to study the secondary structure of proteins by analysing their characteristic β-sheet and α-helix content. Electron microscopy (EM) is a low-resolution technique where the shape of the proteins can be visualised directly (Section 2.1.1). Small-angle scattering (SAS) is a low-resolution technique that includes small-angle X-ray scattering (SAXS) and small-angle neutron scattering (SANS), which can resolve three-dimensional protein structure to a resolution of about 2-4 nm (Perkins *et al.*, 2008). By coupling SAS with constrained homology modelling (Chapter 3) the resolution of structural data obtained by SAS can be increased to medium resolution (about 1-2 nm). Analytical ultracentrifugation (AUC) is another low-resolution technique that obtains hydrodynamic and shape information on the proteins by monitoring their sedimentation behaviour with the centrifugal forces that are applied (Section 2.3). High-resolution protein structure determination methods include nuclear magnetic resonance (NMR) (Section 2.1.2) and X-ray crystallography (Section 2.1.3), both of which resolve protein structures to give structural information at an atomic coordinate level. All of these techniques are often complemented by biochemical and genetic information about the protein, particularly information on the evolution of the proteins and protein homologues that have similar structures or functions. Also, all of these techniques have pros and cons for the structural determination of different proteins. The technique that is best suited to gain the required structural information at as high a resolution as possible is employed.

An antibody (Ab) is an important effector molecule of the humoral immune system (Section 1.2). Determination of the structure of Abs is essential to further understand their specific functions in immunity and their roles in

immune diseases. However, Abs are large, flexible and generally highly glycosylated molecules that can form large polymeric molecules structures or complexes, which makes determination of their structure by EM, NMR or crystallography very difficult. In a situation such as this, SAS and AUC combined with constrained modelling, allows the examination of Abs in near-physiological conditions to result in an Ab solution structure of medium resolution. By this approach, important structural information is revealed, which paves the way for a better understanding of the biological roles of Abs in the human immune system.

2.1.1 Electron microscopy

EM is a similar technique to light microscopy, yet uses an electron beam in order to visualise the shape of protein molecules. EM has a greater magnification and resolving power than light microscopy, yet the high level of radiation can potentially damage the biological sample. EM provides low-resolution structural details, and a low signal to noise ratio means that the information from many macromolecules must be combined and averaged. In a typical solution, protein molecules are not ordered but may exist in several conformations or views. This is particularly true for Ab samples, where the hinge regions and tailpieces increase molecular flexibility. Protein samples are visualised *in vacuo*. It is often necessary to process the protein samples to successfully visualise the protein structure. Protein fixation and staining techniques can damage or distort the protein structure, altering its three-dimensional arrangement from its native conformation in physiological conditions. Determinations of the site of interaction between proteins or the structures of multiple proteins within a complex are difficult to determine by EM and this is exemplified by the two EM structures for SIgA (Figure 1.9; Svehag & Bloth, 1970; Bloth & Svehag, 1971).

2.1.2 Nuclear Magnetic Resonance

NMR can provide information on the topology, dynamics and three-dimensional structure of molecules in solution by utilising the magnetic properties of certain nuclei: ^{1}H, ^{13}C and ^{15}N. NMR experiments apply magnetic fields at different frequencies to labelled or non-labelled protein samples, and monitor the characteristic absorption energy of the nuclei, which may differ in accordance with their chemical microenvironment within the protein. The chemical shift is determined for each nucleus, which corresponds to the environment of the nucleus. Multinuclear NMR permits the assignment of chemical shifts to specific nuclei, and nuclear Overhauser effects are used to determine spatial restraints. This information on the nuclei allows the position of atoms to be assigned and models to be created for the three-dimensional structure of the protein. NMR experiments study proteins in solution, and therefore these are often in near-physiological conditions. NMR is a good option to determine protein structure if crystals cannot be formed for X-ray crystallography (Section 2.1.3). However, structural information cannot be obtained for intact antibodies by NMR. Antibodies have a minimum molecular mass of approximately 180 kDa (and up to 900 kDa for polymeric Abs) and NMR is most suitable for the study of proteins with a maximum molecular mass of 30 kDa.

2.1.3 X-ray crystallography

X-ray crystallography is a technique that allows the individual atoms to be resolved by interpreting the diffraction of X-rays from the repeating unit cells that constitute the crystal lattice. X-ray crystallography can give electron density maps at high structural resolution (0.1-0.3 nm). Protein crystals are grown, and exposed to a monochromatic, collimated X-ray beam. X-rays are scattered by the electrons within the protein, and the diffracted waves are recorded by a detector. The diffraction pattern depends on the phase, amplitude and frequency of the scattered waves. Only those waves that are in phase and constructively interfere with each other produce a positive beam and subsequently a diffraction spot (Rhodes, 2000).

The amplitude and frequency of each wave is known from the diffraction pattern spots, yet the phase information is lost. This is the phase problem in crystallographic experiments. In order to overcome the phase problem to

calculate the phase of the waves, the position of the atoms in the crystal lattice and therefore create a map of the electron density, two methods can be used. If a homologous protein is known and its three-dimensional structure available, a homology model can be created (Section 3.2). The homology model is used to carry out molecular replacement. The phase of the resulting molecular replacement model is calculated and the model is fitted to the electron density map. Manual fitting and computational refinement then gives a final three-dimensional structure at atomic resolution. Alternatively, the multiple isomorphous replacement method involves soaking the protein crystals in solutions of heavy ions such as Hg, Pt or Au, or replacing a few atoms within the protein residues with heavy atoms, e.g. seleno-methionine. The strong scattering of these heavy atoms alters the observed diffraction pattern of the protein, and from this the phase can be determined. The protein atoms are fitted onto the electron density map by trial-and-error, and refined. Refinement methods calculate the *R* factor and *R* free values, which give the goodness-of-fit of the model to the electron density map (Rhodes, 2000).

Many antibody fragments have been crystallised and their three-dimensional structures determined to atomic resolution. However, intact Abs are much more difficult to crystallise. The movement between the antibody Fc and Fab regions, however minimal, make growing a crystal lattice of identical repeating molecules hard to achieve. Also, the carbohydrates commonly present on Abs can increase the solubility of the molecules and prevent crystal formation (Perkins & Bonner, 2008).

2.2 Solution scattering

SAS is a diffraction technique that is used to investigate the overall structure of macromolecules in random orientations in solution at low-resolution (Glatter & Kratky, 1982; Perkins, 1988). The major strengths of solution scattering are that it provides multi-parameter structural and compositional information on proteins that are analysed in near-physiological conditions. Medium-resolution protein models are produced based on the scattering data and known atomic structures (Perkins *et al.*, 2008). SAS can therefore circumvent issues caused by the inability to crystallise the protein of interest.

2.2.1 Complementarity of solution scattering with other methods

SAS can be used to complement other protein structure determination methods. EM directly visualises macromolecules, but this occurs *in vacuo* and the proteins are subjected to processing that can be harsh and potentially alter their structure. Macromolecular dimensions are determined by SAS and these can be used to confirm those obtained by EM.

NMR does give atomic resolution structures in solution, yet it is limited to structures for small proteins with a molecular mass less than about 30 kDa. Scattering is a useful structure determination method for proteins with a molecular mass of 10 kDa or upward, with no upper limit. This means large molecules like Abs or macromolecular complexes can be studied by SAS.

X-ray crystallography results in atomic resolution electron density maps. The outcome can potentially be affected by artefacts caused by intermolecular contacts in the crystal or by conditions used to form the crystal such as high salt. SAS used in conjunction with X-ray crystallography can identify any significant conformational or oligomeric differences between the solution and crystal structure. If the protein is too large, flexible or heavily glycosylated to be crystallised intact, such as antibodies, and atomic structures are available for the constituent domains, SAS provides a medium-resolution structure when accompanied with constrained modelling.

AUC provides sedimentation or diffusion coefficients as a measurement of macromolecular shape and can reveal sample polydispersity. The information obtained by AUC is independent yet complementary to the parameters obtained by SAS, and so both techniques can be used to confirm the findings of the other. As AUC can resolve sedimentation coefficients of individual components in a mixture, this is a powerful supplement to SAS (Section 2.3).

2.2.2 Theory of solution scattering

While electrons diffract X-rays, neutrons are diffracted by nuclei, yet the physical principles are the same in both small-angle X-ray scattering (SAXS) and small-angle neutron scattering (SANS) experiments.

A protein sample is irradiated with a collimated, monochromatic beam of X-ray or neutrons (Figure 2.1a). The scattering of X-rays by protein molecules is due to the interaction of the X-rays with electrons. Neutrons are scattered by the

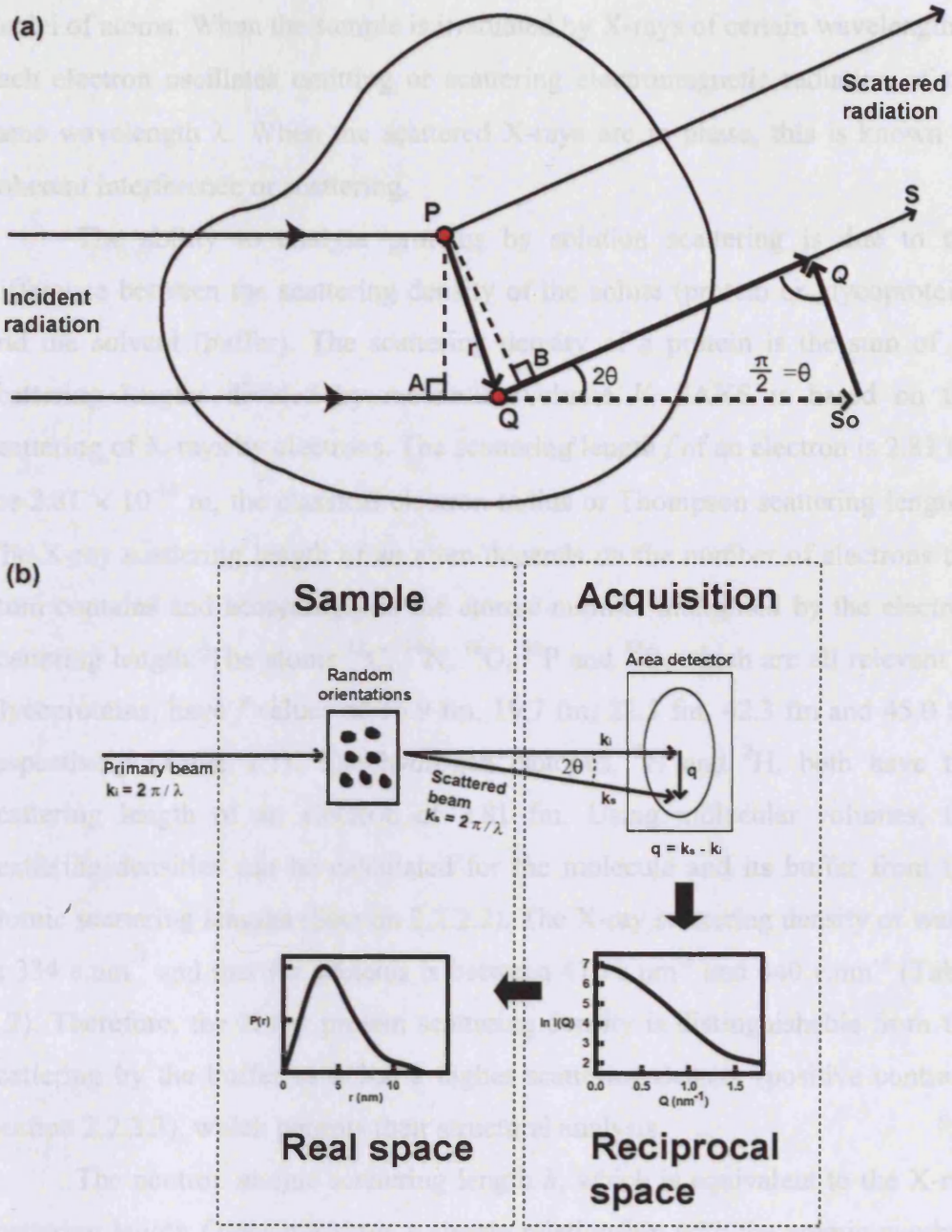

Figure 2.1 Schematic representation of a scattering experiment

(a) Scattering from two points P and Q in a protein molecule are shown (red circles). The scattering angle 2θ and scattering vector (*Q*) are shown. (b) The accumulation of diffraction as shown in (a) gives rise to a scattering pattern on an area detector shown in (b). The radial average of the scattering pattern about the position of the direct main beam (masked by a beam stop) gives rise to the scattering curve *I(Q)* in reciprocal space. Fourier transformation of this to the *P(r)* curve gives real space information on the protein molecule. (Adapted from Pekins *et al.*, 2008).

nuclei of atoms. When the sample is irradiated by X-rays of certain wavelength λ each electron oscillates emitting or scattering electromagnetic radiation of the same wavelength λ. When the scattered X-rays are in phase, this is known as coherent interference or scattering.

The ability to analyse proteins by solution scattering is due to the difference between the scattering density of the solute (protein or glycoprotein) and the solvent (buffer). The scattering density of a protein is the sum of its scattering lengths divided by molecular volume *V*. SAXS is based on the scattering of X-rays by electrons. The scattering length *f* of an electron is 2.81 fm (or 2.81×10^{-15} m, the classical electron radius or Thompson scattering length). The X-ray scattering length of an atom depends on the number of electrons the atom contains and accordingly is the atomic number multiplied by the electron scattering length. The atoms ^{12}C, ^{14}N, ^{16}O, ^{31}P and ^{32}S, which are all relevant to glycoproteins, have *f* values of 16.9 fm, 19.7 fm, 22.5 fm, 42.3 fm and 45.0 fm respectively (Table 2.1). The hydrogen isotopes, ^{1}H and ^{2}H, both have the scattering length of an electron of 2.81 fm. Using molecular volumes, the scattering densities can be calculated for the molecule and its buffer from the atomic scattering lengths (Section 2.2.2.2). The X-ray scattering density of water is 334 $e.nm^{-3}$ and that for proteins is between 410 $e.nm^{-3}$ and 440 $e.nm^{-3}$ (Table 2.2). Therefore, the X-ray protein scattering density is distinguishable from the scattering by the buffer as it has a higher scattering density (positive contrast; Section 2.2.2.3), which permits their structural analysis.

The neutron atomic scattering length *b*, which is equivalent to the X-ray scattering length *f* does not show a simple relationship with the atomic number. SANS differs from SAXS in that neutrons interact directly with the nucleus of the atom, so the scattering length and therefore the contribution to the scattered intensity is different for the ^{1}H and ^{2}H isotopes. The atoms ^{2}H, ^{12}C, ^{14}N, ^{16}O, ^{31}P and ^{32}S have comparable positive *b* values of 6.671, 6.651, 9.400, 5.804, 5.1 and 2.847 fm respectively. However, the ^{1}H isotope has a large, negative *b* value of -3.742 fm (Table 2.1). From the molecular volume, the neutron protein scattering density in $^{2}H_2O$ becomes 40-45% $^{2}H_2O$, based on a scale ranging from 0% to 100% $^{2}H_2O$ (Table 2.2). The large difference in the neutron scattering properties

Atom		Atomic Number	*f* (fm)	*b* (fm)
Hydrogen	^{1}H	1	2.81	-3.742
	^{2}H	1	2.81	6.671
Carbon	^{12}C	6	16.9	6.651
Nitrogen	^{14}N	7	19.7	9.4
Oxygen	^{16}O	8	22.5	5.8
Phosphorous	^{31}P	15	42.3	5.1
Sulphur	^{32}S	16	45.0	2.847

Table 2.1 Scattering lengths *f* (X-rays) and *b* (neutrons) of biologically important atoms

(Taken from Perkins *et al.*, 2008).

	X-rays ($e.nm^{-3}$)	Neutrons (% $^{2}H_2O$)
H_2O	334	0
$^{2}H_2O$	334	100
Lipids	310-340	13
Detergents	300-430	6-23
Proteins	410-440	40-45
Carbohydrates	490	47
DNA	590	65
RNA	600	72

Table 2.2 Scattering densities of solvents and biological macromolecules

(Adapted from Perkins, 1988).

of the two hydrogen isotopes can be utilised in contrast variation neutron scattering experiments (Table 2.2; Section 2.2.2.3).

In SAXS, the effect of incoherent scattering (where the X-ray scattered by the electrons are out of phase) is negligible. In neutron scattering, incoherent scattering exists for nuclei with spin. This is significant for ^{1}H nuclei, where incoherent scattering is very large and leads to a background level of uniformly scattered neutrons, when the wavelength λ is greater than 1 nm. When modelling neutron curves of proteins that were analysed in $^{2}H_2O$ buffers, intensities at large Q can be affected by small background incoherent scattering so a flat-background correction is applied to remove it.

As a result of coherent scattering and the differences in scattering densities between the protein and buffer (the contrast), an intense circularly symmetric diffraction pattern is observed on a two-dimensional area detector at low scattering angles Q (Figure 2.1b). The intensity of scattering $I(Q)$ diminishes as Q increases. The radially-averaged intensities $I(Q)$ are measures as a function of Q, where $Q = 4\pi \sin\theta / \lambda$ (λ = wavelength; 2θ = scattering angle). From this a one-dimensional wide-angle scattering curve $I(Q)$ against Q is obtained (Figure 2.1b; Figure 2.2). The scattering data is obtained in reciprocal space, in a Q range of approximately 0.05 - 2 nm^{-1}. The resolution of a scattering experiment is defined by $2\pi/Q_{max}$ where Q_{max} is the maximum scattering vector measured in the experiment. Classical SAS experiments view structures at a low resolution, of around 3 nm. The scattering curve is truncated at very low Q values due to the presence of the beamstop, which protects the detector and camera from the highest intensity of the direct main beam. At large Q values the scattering curve is truncated by the edge of the detector.

2.2.2.1 The Debye equation

The scattering pattern obtained by SAS is affected by the properties of the biological macromolecule and the spatial arrangement of scattering elements within the macromolecule (Perkins, 1988). The Debye equation considers both of these properties to calculate the $I(Q)$ scattering curve of the macromolecule.

The scattering for any pair of scattering elements (electrons for X-rays and nuclei for neutrons) is shown in Figure 2.1a. The incident X-ray beam is

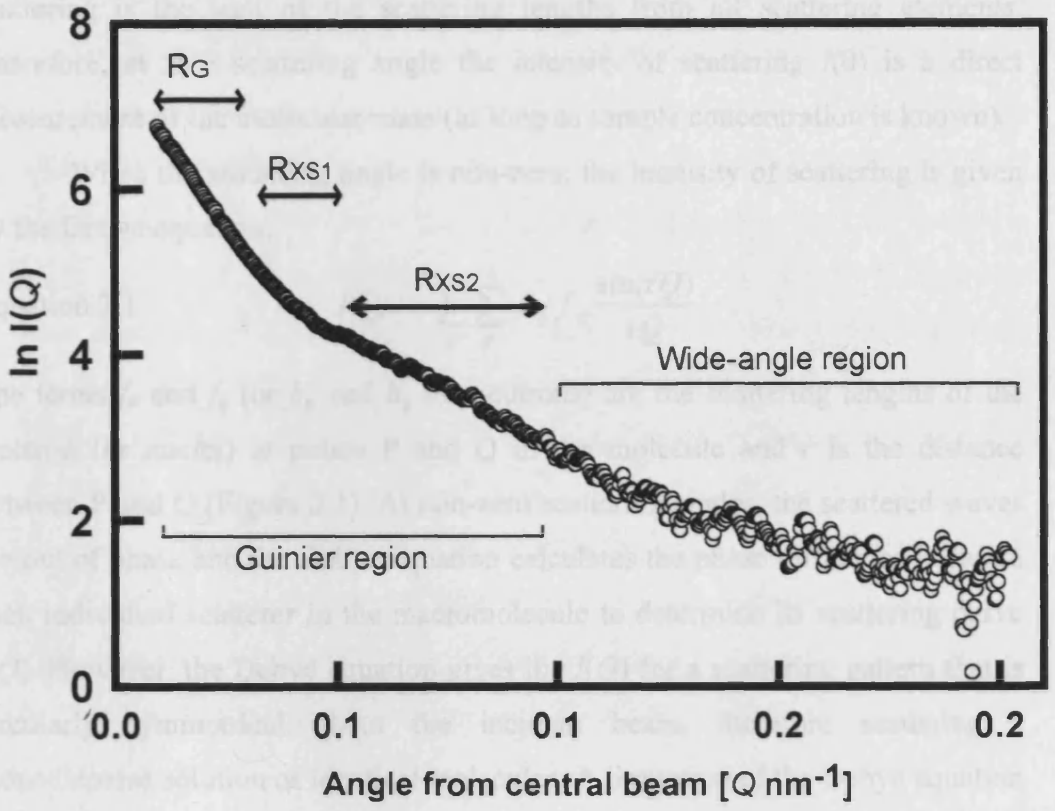

Figure 2.2 General features of an X-ray solution scattering curve

The scattering curve is analysed in two regions. Analysis of low Q regions gives the Guinier plots and the forward scattering intensity $I(0)$ value. Analysis of wide-angle region at larger Q gives more structural information. At low Q the scattering curve is truncated for reason of the beamstop. The scattering curve shown is that obtained for dIgA1 (Chapter 5).

defined by the unit vector s_0 and is scattered from a scattering element in a direction denoted by the unit vector s and at scattering angle 2θ. Q is the scattering vector $(s - s_0)$, and its amplitude is $4\pi \sin\theta/\lambda$. When 2θ is zero the scattered waves scattered from all points are in phase and the intensity of scattering is the sum of the scattering lengths from all scattering elements. Therefore, at zero scattering angle the intensity of scattering $I(0)$ is a direct measurement of the molecular mass (as long as sample concentration is known).

When the scattering angle is non-zero, the intensity of scattering is given by the Debye equation:

Equation 2.1
$$I(Q) = \sum_p \sum_q f_p f_q \frac{\sin(rQ)}{rQ}$$

The terms f_p and f_q (or b_p and b_q for neutrons) are the scattering lengths of the electron (or nuclei) at points P and Q in the molecule and r is the distance between P and Q (Figure 2.1). At non-zero scattering angles, the scattered waves are out of phase and the Debye equation calculates the phase difference between each individual scatterer in the macromolecule to determine its scattering curve $I(Q)$. However, the Debye equation gives the $I(Q)$ for a scattering pattern that is circularly symmetrical about the incident beam, therefore assuming a monodisperse solution of identical molecules. A limitation of the Debye equation is that it describes particles *in vacuo*. Therefore, the Debye equation is modified in order to consider a two-phase model of aqueous proteins in a solution.

2.2.2.2 The two-phase model of solution scattering

An aqueous protein solution can be considered a two-phase system consisting of solute and solvent. The Debye equation is modified to consider the contrast $\Delta\rho$, which is the difference between the mean scattering density of the macromolecule ρ_v and that of its buffer ρ_s:

Equation 2.2
$$\Delta\rho = \rho_v - \rho_s.$$

The terms ρ_v and ρ_s are calculated from the total scattering lengths divided by the volume V and replace the atomic scattering lengths Σf_p or Σb_p (Equation 2.1; Perkins, 1988). A compilation of the scattering densities of some solvents and macromolecules are given in Table 2.2.

2.2.2.3 Contrast variation

In biology, the variations in the scattering densities of different macromolecules are very useful in structural analysis of complexes or internal structures (Table 2.2), for example, protein-DNA or protein-lipid structures.

The main difference between scattering densities of the macromolecular classes by SAXS and SANS is the wider range of solvent contrast variation in the SANS. Neutron contrast variation using H_2O-^2H_2O encompasses the scattering densities of all macromolecular classes. Lipids, proteins, carbohydrates, DNA and RNA each have their own characteristic scattering density that reflect the different density of the proton in each one so that SANS can differentiate between them. They all have scattering densities between that of H_2O and ^2H_2O (Table 2.2). Therefore, a particular ratio of H_2O and ^2H_2O in buffers or deuteration of a component in a complex can be carefully chosen to select the contrast that will reveal the most structural information of the macromolecule. Due to the low amounts of non-protein macromolecules studied in this thesis, the neutron work presented here did not involve contrast variation experiments. Instead, proteins were studied in the highest solute-solvent neutron contrasts, i.e. 100% ^2H_2O, in order to maximise the strength of the scattered radiation. X-rays study macromolecules in high positive solute-solvent contrast, which is the opposite of the high negative solute-solvent contrast of neutrons (Perkins *et al.*, 2008). Analysis of both provides a control of the inhomogeneity effects caused by the presence of large amounts of carbohydrate moieties.

2.2.2.4 The hydration shell

The neutron scattering properties of ^1H atoms necessitates that neutron experiments are performed in 100% ^2H_2O buffers in order to minimise incoherent scattering and obtain scattering curves comparable to those measured in X-ray experiments. Even under these conditions, however, small differences are often observed between measurements from X-ray and neutron experiments, and these are attributed to the hydration shell that surrounds protein molecules in solution. The hydration shell refers to those solvent water molecules closely associated with the protein by way of hydrogen bonding. While in X-ray scattering experiments the hydration shell is visible, in neutron scattering experiments the hydration shell is present yet it is almost undetectable. The

model of scattering by an aqueous protein solution assumes that the solute and solvent are discrete entities, although in reality water molecules are hydrogen bonded to the protein surface. A water molecule that is bonded to the surface of a protein is electrostricted and therefore has a smaller volume (0.0245 nm^3) than a water molecule in the bulk solvent (0.0299 nm^3) (Perkins, 1986). As a result, the X-ray scattering density of water molecules hydrogen-bonded to the surface of a protein is increased relative to that of a water molecule in the bulk solvent, becoming more similar to the X-ray scattering density of the protein. X-rays typically detect a hydration shell of 0.3 g H_2O for each gram of protein (Perkins, 1986). In neutron experiments the hydration shell is not visible due to the rapid exchange of ^1H and ^2H atoms between the hydration shell and the bulk solvent, giving rise to little difference in scattering densities between the two.

2.2.3 Solution scattering methods

2.2.3.1 Sample preparation

X-ray solution scattering experiments are designed according to the two-phase model of solution scattering, with a monodisperse protein solution in a buffer (Section 2.2.2.2). The choice of buffer is important since in an X-ray scattering experiment, the greater the difference in scattering densities between the buffer and protein, the higher the contrast and sample transmission, and therefore the better the counting statistics. X-rays can be absorbed by high-salt buffers, which if used can increase the background scattering. Phosphate buffered saline (116 mM NaCl, 10.4 mM Na_2HPO_4, 3.2 mM NaH_2PO_4 at pH 7.2) was used for X-ray scattering experiments. In order to ensure that the appropriate corrections are made for solvent scattering, the protein samples were dialysed against their buffer shortly before data collection, and the scattering of the buffer was then subtracted from the scattering of the protein solution. Ideally, a sample volume of at least 1 ml with a maximum concentration above 2 mg/ml was used for X-ray experiments. A concentration series, with at least four concentrations was used to measure the protein to ensure consistent results and that no concentration-dependence occurred. The concentration of a protein solution can be determined from its tryptophan, tyrosine and cysteine content by measuring its absorbance at 280 nm, and using an absorption extinction coefficient calculated from its amino acid and carbohydrate composition (Perkins, 1986; Section 3.3.1).

The optical densities were measured in a UV spectrophotometer in quartz cells with known pathlengths. The concentration is calculated from the Beer-Lambert law, $c = A / \varepsilon l$, where c is the protein concentration (mg/ml), A is absorbance at wavelength 280 nm, ε is the protein extinction coefficient (mg/ml/mm) and l is the pathlength (mm).

Samples for neutron scattering experiments were prepared similarly as for X-ray scattering experiments, except that PBS buffer in 100% ^2H_2O was used. Each protein sample was dialysed with stirring against the ^2H_2O PBS buffer for a minimum of 36 hours, in which time the buffer was changed four times to ensure full ^1H-^2H exchange.

2.2.3.2 X-ray data collection on instrument ID02 at the ESRF

X-ray scattering experiments were performed at the European Synchrotron Radiation Facility (ESRF) at Grenoble, France (Figure 2.3a,b). Synchrotron radiation is emitted from electrons that are accelerated while moving at speeds close to the speed of light. Several stages are used in the production of synchrotron radiation at the ESRF. Electrons are produced by a triode gun and then accelerated to 200 MeV in a linear accelerator (LINAC). The electrons are then transferred to a 300m circumference booster synchrotron where they are accelerated to 6.0 GeV. The electrons are then injected into the storage ring. The purpose of the 844.4 m diameter storage ring is to store the 6.0 GeV electron beam after injection from the booster. Sixty-four bending magnets (BMs) are employed to force the electrons to follow a pseudo-circular orbit (Figure 2.3b). Insertion devices (IDs) are used to force the beam onto a sinusoidal trajectory, stimulating additional radiation (Figure 2.3b). As the electrons pass through the BMs and IDs, they lose energy and so radio-frequency cavities are used to maintain the energy at 6 GeV. The intensity and lifetime (the time the electrons can be kept in orbit for before the storage ring has to be refilled) of the beam depends on the operation of the storage ring. There are two modes of operation. Multibunch operation involves a long pulse (1 μs) of the LINAC delivering 992 electron bunches to the booster ring, filling its circumference, where in the storage ring the electron bunches are equally distributed around its circumference. This has a beam intensity of 200 mA and a

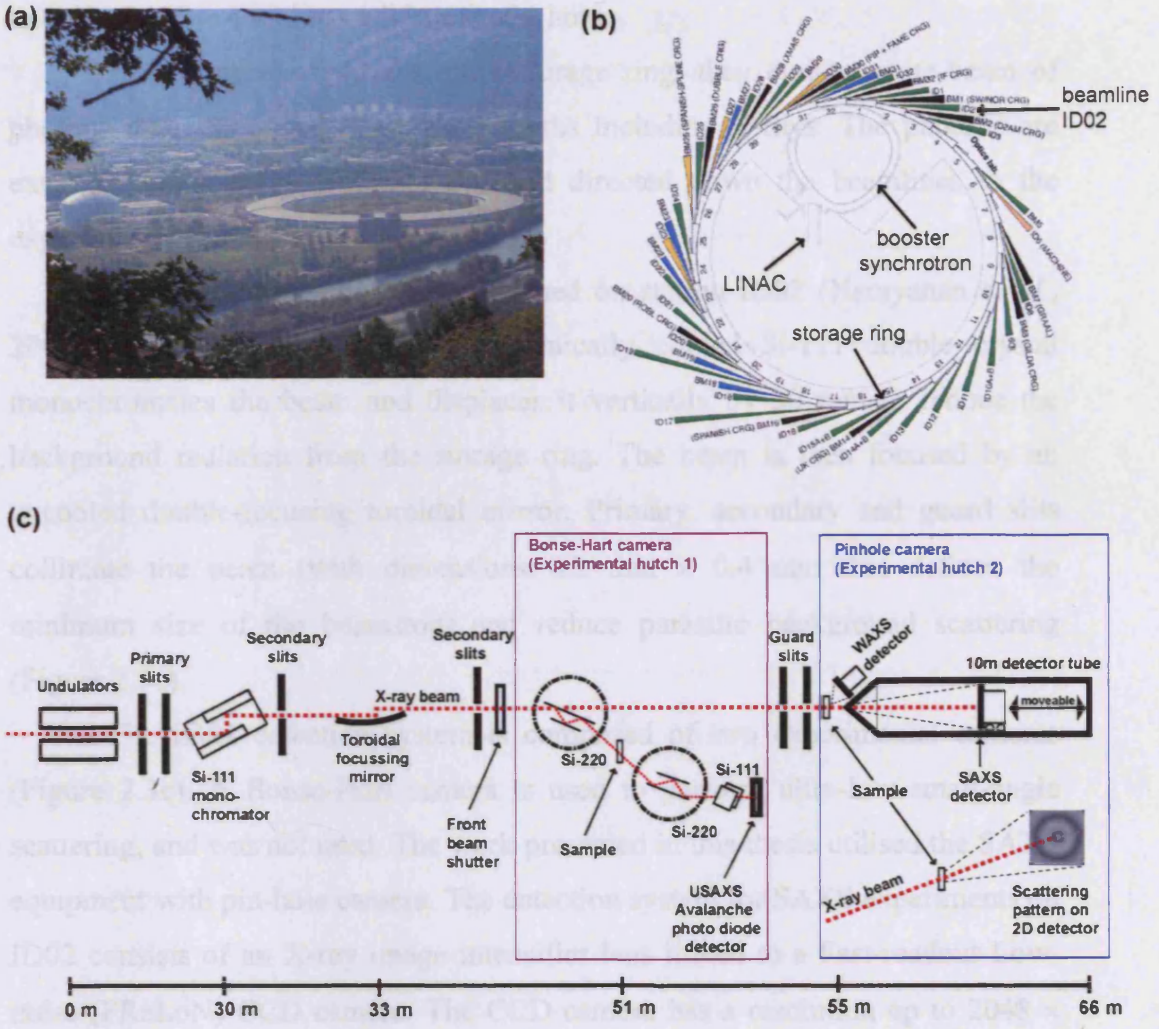

Figure 2.3 X-ray solution scattering at the ESRF

(a) Photograph of the ESRF synchrotron ring in Grenoble. (b) Schematic representation of the synchrotron. The linear accelerator (LINAC), booster synchrotron and the storage ring are shown. The SAXS beamline ID02 is highlighted. (b) Schematic representation of the layout of the X-ray solution scattering cameras at beamline ID02. (Adapted from the ESRF Web Site *http://www.esrf.fr*).

lifetime of approximately 60 hours. The second mode of operation is single-bunch operation, where a short pulse (2 ns) of the LINAC only delivers one electron bunch to the booster and subsequently the storage ring, and the beam intensity is 16 mA and has a lifetime of 6 hours.

As the electrons circulate the storage ring, they emit a white-beam of photons that comprises all wavelengths including X-rays. The photons are extracted tangentially to the source and directed down the beamlines to the experimental stations (Figure 2.3b).

SAXS experiments were performed on station ID02 (Narayanan *et al.*, 2001; Panine *et al.*, 2006). A cryogenically cooled Si-111 double crystal monochromates the beam and displaces it vertically by 30 mm to reduce the background radiation from the storage ring. The beam is then focused by an uncooled double-focusing toroidal mirror. Primary, secondary and guard slits collimate the beam (with dimensions 0.2 mm × 0.4 mm that defines the minimum size of the beamstop) and reduce parasitic background scattering (Figure 2.3c).

The ID02 detection system is composed of two experimental cameras (Figure 2.3d). A Bonse-Hart camera is used to perform ultra-low small-angle scattering, and was not used. The work presented in this thesis utilised the SAXS equipment with pin-hole camera. The detection system for SAXS experiments on ID02 consists of an X-ray image intensifier lens linked to a Fast-readout Low-noise (FReLoN) CCD camera. The CCD camera has a resolution up to 2048 × 2048 pixels. A dark image (with no X-ray beam exposure) and a raw scattered image are measured, where the dark image is then subtracted from the raw image. The detector is housed in a vacuum within a long detector tube on a movable platform so that the sample-to-detector distance can be varied between 1 and 10 m. For the X-ray work within this thesis, the distance between the detector and the sample was set to 3 m so that intensities were measured to a maximum Q value of 2.5 nm^{-1}. A Lupolen sample is used for the calibration of detector intensities (Kratky *et al.*, 1966), and a silver salt standard (silver behenate) is used for the calibration of the Q range (Huang *et al*, 1993). The X-ray wavelength on ID02 is 0.1 nm.

The protein samples were placed in a specially designed Perspex cell with 10 to 20 μm thick ruby mica windows. The mica windows were held in place with Teflon plugs. The sample cell had a 1 mm path length, a surface area of 2 mm (vertical) by 8 mm (horizontal) and held a maximum volume of 25 μl. The sample cell was positioned in the X-ray beam and held in place by a brass sample holder which was maintained at 15°C using a water bath. Prior to data collection, the sample holder was manually aligned in the beam.

X-rays produce free radicals that can cause proteins to aggregate quickly. In order to monitor possible protein aggregation, the scattering intensity of each sample was measured for a set time in 10 equal time frames, and the first and last frames were inspected to check for radiation damage. The ID02 SAXS detector is interfaced with a computer for data accumulation, which is networked to a workstation for data storage and processing. The online data reduction meant that the collected data was converted to the one-dimensional scattering curve that could be almost instantly visually inspected during data acquisition.

The protein samples were measured in alternation with their respective buffers in order to minimise errors in the buffer subtraction calculations as the incident beam decreased in intensity.

2.2.3.3 Neutron data collection on Instrument LOQ at ISIS

SANS experiments were performed on the LOQ instrument at the pulsed neutron source at ISIS at the Rutherford Appleton Laboratory (RAL) in Didcot, U.K. (Heenan *et al.*, 1997). The neutrons at ISIS are produced by a spallation process (Figure 2.4a) that is designed to accommodate the pulsed main neutron beam.

High-energy H^- ions pass though a LINAC where protons are then produced by stripping the electrons from the H^- ions using a very thin aluminium foil. Pulses of protons are then accelerated in a 52 m diameter synchrotron, extracted and sent to the heavy metal target station (Figure 2.4a). Uranium (U) or tantalum (Ta) heavy metal targets are bombarded with high-energy protons, causing neutrons to be 'chipped' out of the nuclei. Moderators are used to slow the neutrons and increase the wavelengths. On LOQ the moderator is liquid hydrogen, positioned 11.1 m from the sample rack. A supermirror bender removes neutrons with wavelengths of less than 0.2 nm, while another mirror

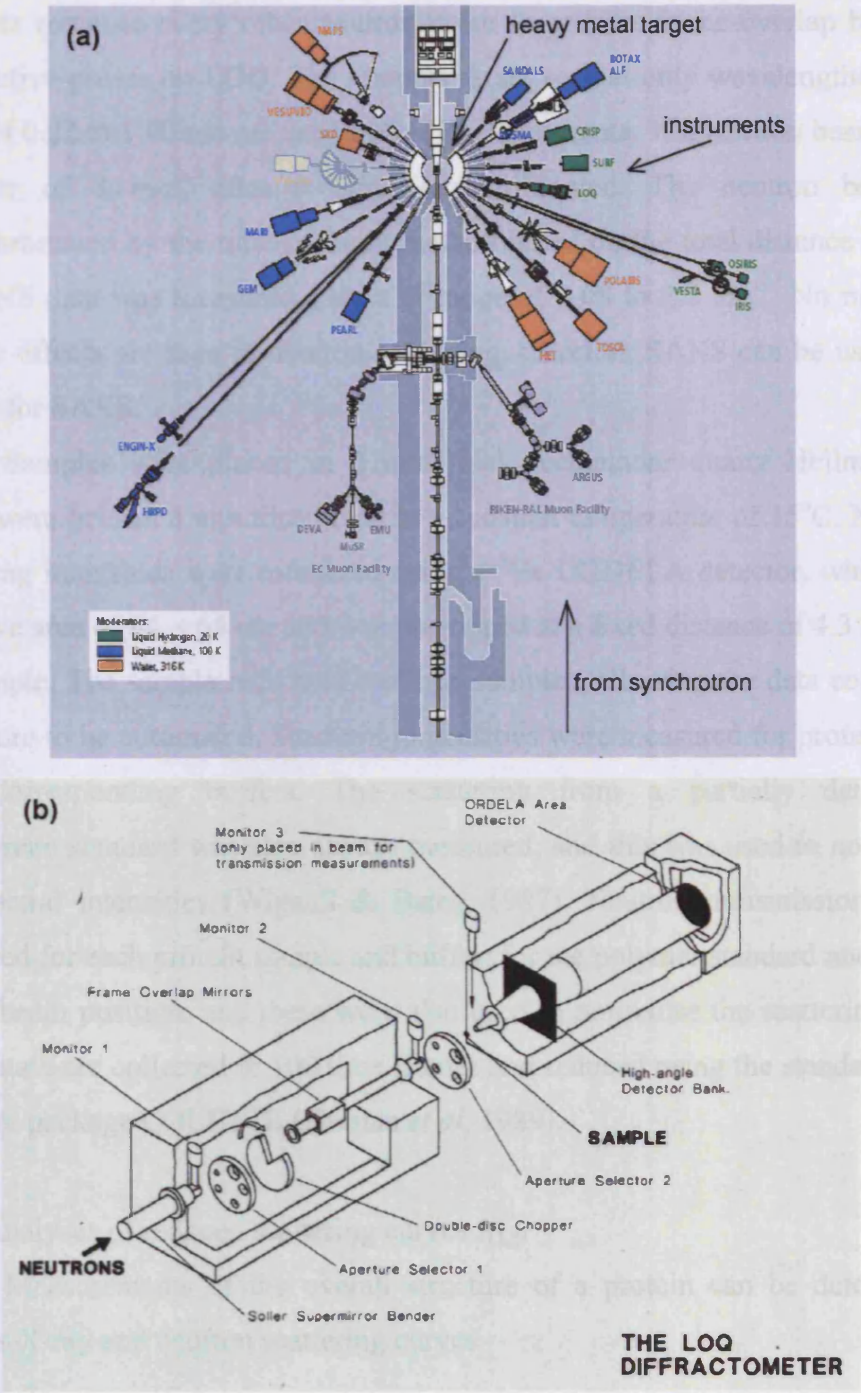

Figure 2.4 Neutron solution scattering at ISIS.

(a) Layout of the target station at ISIS located at the Rutherford-Appleton Laboratory, Didcot (Adapted from the ISIS Web Site *http://www.isis.rl.ac.uk*). (b) Schematic diagram of the LOQ diffractometer. (Taken from the Rutherford-Appleton Laboratory Web Site *http://www.rl.ac.uk*).

removes those with wavelengths of greater than 1.2 nm. A disc chopper spinning at 25 Hz removes every other neutron pulse to prevent frame-overlap between consecutive pulses on LOQ. The chopper is set so that only wavelengths in the range of 0.22 to 1.00 nm are used in SANS experiments. The neutron beam has a diameter of 8 mm, after it has been collimated. The neutron beam is monochromated by the time-of-flight method based on the total distance of 15.5 m. SANS data was measured over a Q range of 0.08 to 2.5 nm^{-1}. No radiation damage effects are seen in neutron scattering, therefore SANS can be used as a control for SAXS.

Samples were placed in 2 mm thick rectangular quartz Hellma cells which were held in a motorised rack at a constant temperature of 15°C. Neutron scattering intensities were measured using a ^{3}He ORDELA detector, which had an active area of 64 × 64 cm and was positioned at a fixed distance of 4.3 m from the sample. The sample rack held multiple samples, allowing the data collection procedure to be automated. Scattering intensities were measured for proteins and their corresponding buffers. The scattering from a partially deuterated polystyrene standard was also always measured, and this was used to normalise the spectral intensities (Wignall & Bates, 1987). Neutron transmissions were measured for each protein sample and buffer, for the polymer standard and for an empty beam position, and these were also used to normalise the scattering data. Raw data were collected in 100 time frames and reduced using the standard ISIS software package COLETTE (Heenan *et al*, 1989).

2.2.4 Analyses of reduced scattering curves *I(Q)*

Measurements of the overall structure of a protein can be determined from its X-ray and neutron scattering curves.

2.2.4.1 Guinier analyses

At low Q values, the Debye equation (Equation 2.1) becomes the Guinier approximation (Guinier & Fournet, 1955):

Equation 2.3 $$\ln I(Q) = \ln I(0) - \frac{R_G^2 Q^2}{3}$$

where $I(0)$ is the intensity at zero Q, and R_G is the radius of gyration which is defined as the root-mean-square distance of all scattering centres in the

macromolecule from its centre of gravity. The R_G characterises the degree of elongation of the protein, and is analogous to the sedimentation coefficient from AUC experiments (Section 2.3). Guinier plots are only valid at low Q, in a $Q.R_G$ range of up to 0.7-1.3 depending on the macromolecular shape. Intensities at the lowest Q cannot be measured due to the presence of the beamstop. R_G and $I(0)$ are determined from an experimental scattering curve by plotting ln $I(Q)$ against Q^2 to give a straight line of slope $-R_G^2/3$ and intercept ln $I(0)$.

A useful application of R_G is the calculation of the anisotropy ratio of a macromolecule. The anisotropy ratio is R_G/R_0, where R_0 is the R_G of a sphere of equal volume to the macromolecule. For typical globular proteins by X-ray scattering, the anisotropy ratio is approximately 1.28 (Perkins, 1988). The anisotropy ratio standardises the R_G value by considering molecular mass (M_r), and therefore permits the comparison of molecular elongation between molecules.

The intensity at zero scattering angle $I(0)$ is proportional to M_r^2 (Figure 2.5). Molecular mass can be determined from $I(0)/c$ either as relative values from SAXS or absolute values from SANS, if the scattering intensities of are normalised to a partially deuterated polystyrene standard. There is a linear relationship between the M_r determined from composition analyses and $I(0)/c$, which is $M_r = I(0)/c \times 9 \times 10^5$ (Boehm *et al.*, 1999) and can confirm molecular mass of proteins and hence their oligomeric state (Figure 2.5). The M_r calculation validates the Guinier analyses.

2.2.4.2 The cross-sectional radius of gyration

If the protein has an elongated structure, Guinier-type analyses of the scattering curve at larger Q values will give the mean radius of gyration of the cross-section R_{XS}:

Equation 2.4 $$\ln[I(Q)Q] = \ln[(I(Q)Q)]_{Q\to 0} - \frac{R_{XS}^2 Q^2}{2}$$

R_{XS} is determined from an experimental scattering curve by plotting $\ln[I(Q)Q]$ against Q^2 to give a straight line of slope $-R_{XS}^2/2$ and intercept $\ln[(I(Q)Q)]_{Q\to 0}$. The Q ranges to determine R_{XS} do not overlap with that used for the determination of R_G (Figure 2.2).

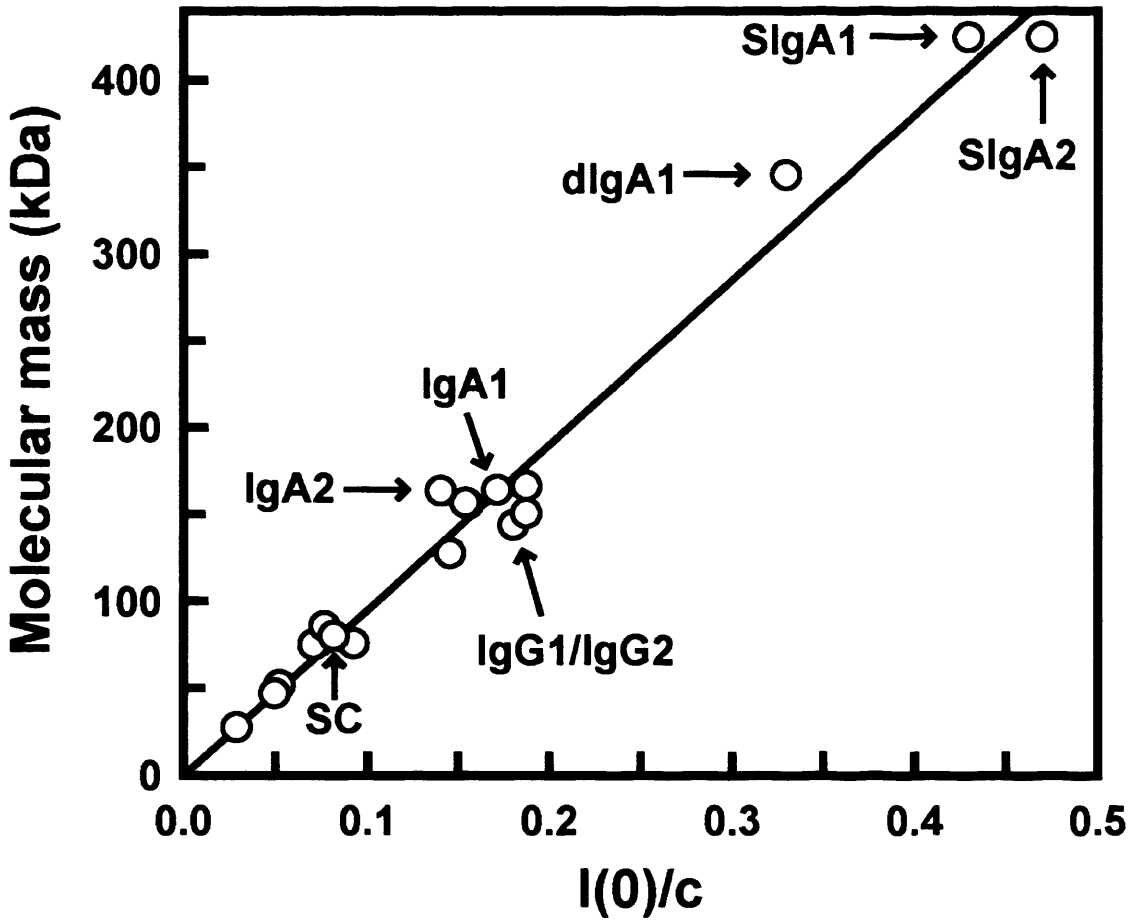

Figure 2.5 Molecular mass against *I(0)/c* for proteins measured on LOQ

Linear relationship between molecular mass and the neutron *I(0)/c* is shown, using the values from 18 glycoproteins in 100% ^2H_2O buffers. The data points of IgA1, IgA2, IgG1/IgG2, SC, dIgA, SIgA1 and SIgA2 are highlighted. Comparison of the resulting molecular mass of the scattering sample to the sequence-derived molecular mass enables the oligomeric state of the protein in solution to be determined.

For bent protein structures such as T- or Y- shaped Abs, the cross-sectional plot exhibits two linear regions, a steeper innermost one and a flatter outermost one (Pilz *et al.*, 1973). In this type of analysis, the two regions are referred to as R_{XS-1} and R_{XS-2} respectively (Figure 2.2).

2.2.4.3 R_G and R_{XS} determination of macromolecular dimensions

Assuming a protein is shaped like an elongated elliptical cylinder, the R_G and R_{XS} analyses can be combined to determine its length L (Glatter & Kratky, 1982):

Equation 2.5 $$L = \sqrt{12(R_G^2 - R_{XS}^2)}$$

2.2.4.4 The real space distance distribution function

The scattering curve $I(Q)$ represents the macromolecular structure in reciprocal space. The experimental $I(Q)$ curve can be converted by Fourier transformation into the distance distribution function $P(r)$, which represents the macromolecular structure in real space with units in nm (Figure 2.1b; Perkins *et al.*, 2008). The solution structure of the protein can now be directly visualised as $P(r)$, which corresponds to the distribution of all the distances r between any two elements within the macromolecule. The maximum M in the $P(r)$ curve is the most frequently occurring intermolecular distance. The maximum length of the molecule L is determined from when the $P(r)$ reaches zero at larger values of r. A calculation of R_G is also obtained from the $P(r)$ curve offering an alternative method to the Guinier analysis to confirm the R_G value.

The transformation from reciprocal space into real space is ordinarily unstable since data points are missing at very low Q due to the presence of the beamstop and at high Q at the edge of the detector. Also, high signal to noise ratios are associated with measurements at large Q. Calculation of the $P(r)$ is performed using the GNOM program (Semenyuk & Svergun, 1991) which contains an automated regularisation procedure that stabilises the $P(r)$ calculation and minimises the termination errors. Successful $P(r)$ curves depend on several criteria: (i) $P(r)$ should exhibit positive values; (ii) The R_G from GNOM should agree with that from the Guinier analyses; (iii) $P(r)$ should be zero when r is zero; (iv) $P(r)$ should be stable and reproducible for different experimental $I(Q)$

curves. D_{max} is the maximum macromolecular dimension, and corresponds to the maximum length L. Peaks M in the $P(r)$ curve give the r value of the most frequently occurring distances in the macromolecule and can be informative on the overall shape of the macromolecule. For example, if M is half of L, then the most frequently occurring distance is half of the maximum length and shows that the protein is spherical or globular. Multidomain proteins, including antibodies, sometimes give rise to more peaks M_1 and M_2 at different r values, which suggest they have semi-rigid complex structures.

2.3 Analytical ultracentrifugation

Analytical ultracentrifugation (AUC) is a dynamic technique that uses the effects of centrifugal forces to characterise the structure of proteins in solution. The movement of proteins in a cell is monitored whilst they are subjected to high centrifugal forces.

An analytical ultracentrifuge consists of an optical system and a rotor, and specially designed cells within the rotor (Figure 2.6). At the beginning of an AUC experiment, the solute is present in a monodisperse, uniform concentration throughout the sample cell. As the rotor spins, the distribution of the protein molecules changes as they are subjected to centrifugal forces monitored in real time by the spectroscopic optics system. Two different spectroscopic methods can be used, based on either the absorption or refractive index difference between the sample and the reference buffer cell measured by absorbance optics (Section 2.3.1) and interference optics (Section 2.3.2) respectively. Both spectroscopic methods follow the sedimentation behaviour of the molecules within the cell and can be used in the two types of AUC experiments: sedimentation velocity (Section 2.3.3) and sedimentation equilibrium (Section 2.3.4).

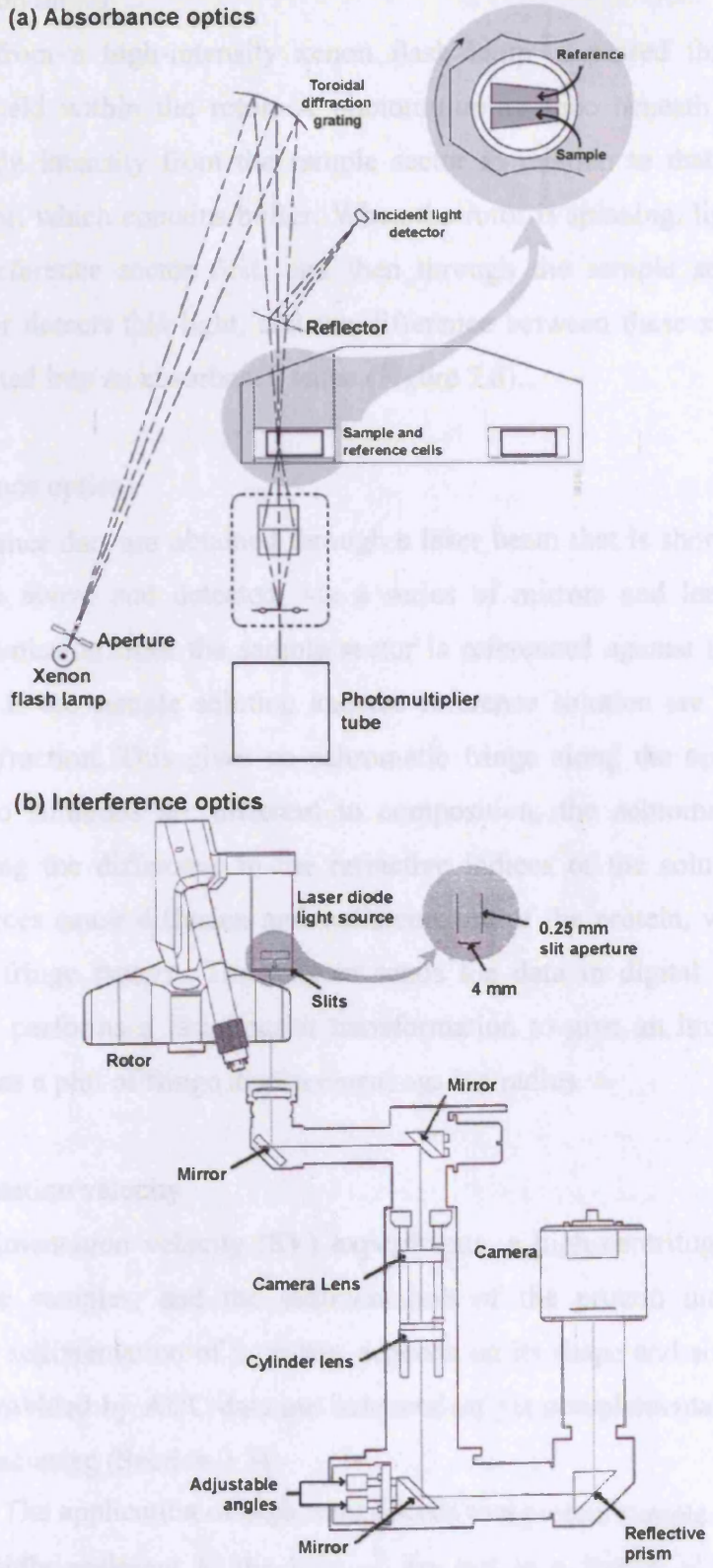

Figure 2.6 Schematic diagram of AUC optic systems

(a) Absorbance and (b) interference AUC optic and rotor systems are shown. (Adapted from Ralston, 1993).

2.3.1 Absorption optics

Light from a high-intensity xenon flash lamp is passed through the sample cells held within the rotor. A photomultiplier tube beneath the rotor detects the light intensity from the sample sector in relation to that from the reference sector, which contains buffer. When the rotor is spinning, light passes through the reference sector first, and then through the sample sector. The photomultiplier detects this light, and the difference between these signals can then be translated into an absorbance value (Figure 2.6).

2.3.2 Interference optics

Interference data are obtained through a laser beam that is shone through the cells from above and detected, via a series of mirrors and lenses, by a camera. The emission from the sample sector is referenced against that of the buffer sector. If the sample solution and the reference solution are the same, there is no refraction. This gives an achromatic fringe along the optical axis. When the two solutions are different in composition, the achromatic fringe shifts, reflecting the difference in the refractive indices of the solutions. The centrifugal forces cause diffusion and sedimentation of the protein, which then displaces the fringe pattern. The camera sends the data in digital form to a computer that performs a fast Fourier transformation to give an image of the fringe pattern as a plot of fringe displacement against radius.

2.3.3 Sedimentation velocity

In sedimentation velocity (SV) experiments, a high centrifugal force is applied to the samples, and the sedimentation of the protein molecules is analysed. The sedimentation of a protein depends on its shape and size, and the information provided by AUC data are independent yet complementary to those obtained by scattering (Section 2.2).

The application of high rotor speeds to a protein sample causes the protein to rapidly sediment to the base of the cell in a matter of hours. SV experiments are performed in long 12 mm columns within the cell. The sedimentation gives rise to the formation of a boundary between the solvent at the meniscus which is depleted of solute and the rest of the protein solution at uniform concentration. The rate of movement of the boundary is measured in real

time, allowing the determination of the sedimentation coefficient, $s^0_{20,w}$ (in standard conditions: in water at 20°C and 1 atmosphere pressure). The units of $s^0_{20,w}$ are seconds, although in ultracentrifugation work they are usually given in Svedbergs, S. One Svedberg is equal to 10^{-13} seconds. The sedimentation coefficient is directly proportional to the mass of the protein, and indirectly proportional to the frictional coefficient f as shown in the simplified Svedberg equation, $S \propto M / f$. Higher S values reflect a higher molecular mass or a more compact protein shape. The frictional coefficient f increases as the protein is more elongated. The frictional ratio f/f_0, where f_0 is the frictional coefficient for a sphere of the same volume as the protein, shows the extent of elongation and is comparable with the scattering anisotropy ratio R_G/R_0.

SV data are collected over 16 hours at rotor speeds between 10,000 r.p.m. and 42,000 r.p.m for protein samples in a dilution series. After correction for the temperature and providing the buffer density ρ and protein partial specific volume $\bar{v}$ are known, the $s^0_{20,w}$ can be obtained by fitting the boundary fits (Figure 2.7a). The buffer density ρ is calculated using the program SEDNTERP (Laue *et al.*, 1992) and the protein partial specific volume $\bar{v}$ is calculated from its amino acid and carbohydrate composition (Section 3.3.1).

Two programs, DCDT+ and SEDFIT can be used to analyse the SV scans to obtain the sedimentation coefficient, after correction for standard conditions. DCDT+ time-derivative analyses subtracts between approximately 4 to 20 pairs of scans and extrapolates the individual subtractions to the start time to give the *g(s*)* function. From this, the sedimentation coefficient, and hence the molecular mass are determined from the peak position and width respectively. The number of pairs of scans must be chosen carefully to ensure time-broadening effects are negligible (Philo, 2000). SEDFIT is a powerful program that carries out direct boundary fits of all the scans simultaneously. Species distribution analyses gives *c(s)* plots (or *c(M)* plots) that show the sedimentation coefficients (or the molecular mass values) of all species within a sample (Figure 2.8; Schuck, 1998). This is a good supplement to the scattering analyses which do not report parameters for different species within a sample. SEDFIT also can calculate the f/f_0 value to confirm that determined from the DCDT+ sedimentation coefficient.

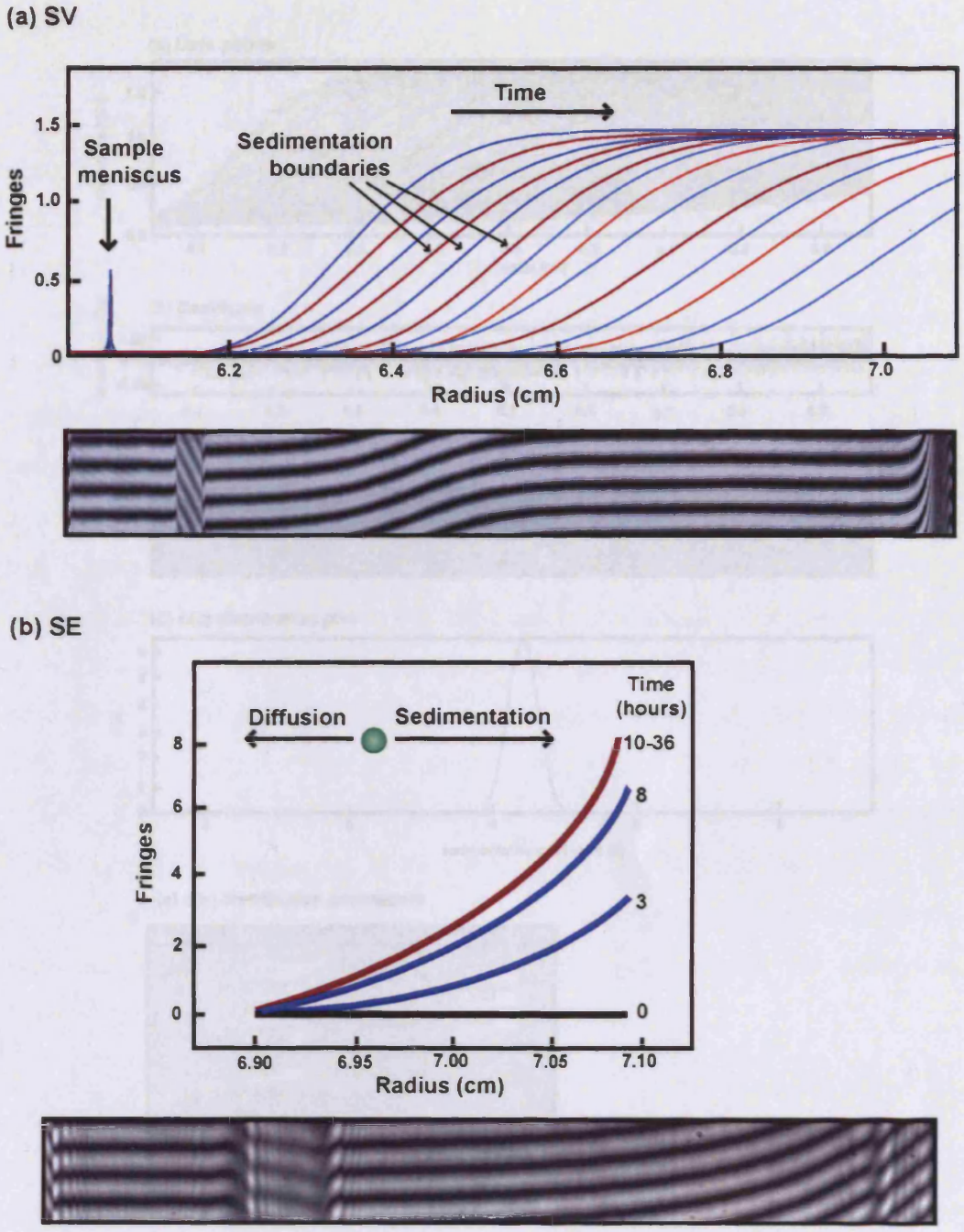

Figure 2.7 Schematic and bitmap image showing SV and SE boundaries

(a) SV data are analysed to obtain the sedimentation coefficient by the fitting of boundaries that represent the sedimentation of the protein in real-time. The sample meniscus is shown. (b) SE data are collected when the protein has reached equilibrium between its diffusion and sedimentation in the cell. The red curve represents the equilibrium state. Below the SV and SE schematic representations are images of the interference fringes recorded during the AUC experiments.

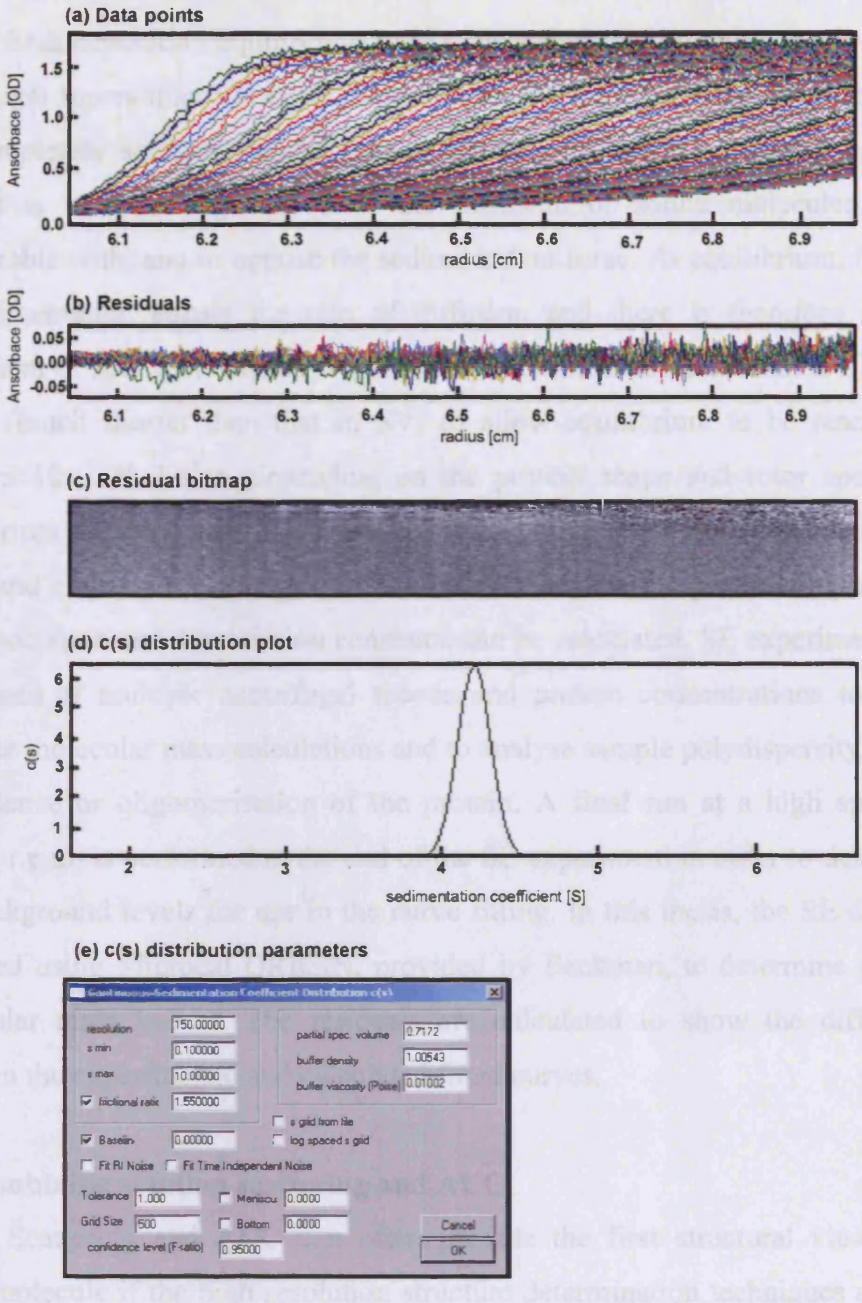

Figure 2.8 SEDFIT analysis output

(a) The SV boundaries and fits to these are represented as data points. (b,c) The goodness-of-fit is reflected in the residual plot and bitmap. A good fit is shown by a monochromatic grey bitmap image and subsequently a low r.m.s.d. value. (d) The resulting fit gives the distribution of sedimentation coefficients of all species in the sample. (e) The parameters window to input the partial specific volume, buffer density and other fit parameters is shown. The SV data shown are those obtained for SC (Chapter 4).

2.3.4 Sedimentation equilibrium

Sedimentation equilibrium (SE) experiments involve much lower centrifugal forces than SV experiments. In SE systems, the solute molecules do not completely sediment to the base of the cell. Instead, the centrifugal force applied is small enough to allow the diffusion of solute molecules to be comparable with, and to oppose the sedimentation force. At equilibrium, the rate of sedimentation equals the rate of diffusion and there is therefore no net movement of solute molecules (Figure 2.7b). The SE sample column is 2 mm in length (much shorter than that in SV) to allow equilibrium to be reached in between 10 – 36 hours, depending on the protein shape and rotor speed. At equilibrium the exponential distribution of the solute is measured and providing that ρ and $\bar{v}$ are known, the molecular mass, the stochiometry of a complex, and any association and dissociation constants can be calculated. SE experiments are performed at multiple centrifugal speeds and protein concentrations to allow accurate molecular mass calculations and to analyse sample polydispersity, speed dependence or oligomerisation of the protein. A final run at a high speed of 42,000 r.p.m. is performed at the end of the SE experiment in order to determine the background levels for use in the curve fitting. In this thesis, the SE data are analysed using Microcal ORIGIN, provided by Beckman, to determine protein molecular mass values. The residuals are calculated to show the difference between the experimental and calculated fitted curves.

2.4 Combining solution scattering and AUC

Scattering and AUC can often provide the first structural view of a macromolecule if the high resolution structure determination techniques are not feasible. Scattering and AUC each provide independent, yet similar structural and compositional information about proteins in solution. The molecular mass (to confirm that calculated from the protein or DNA sequence) can be determined by $I(0)/c$ by scattering and AUC SE experiments. Subsequently, the oligomeric state and any associative properties of the protein can be determined. Both SAS and AUC SV analyse proteins in solution to derive information on their molecular elongation, through the R_G and S respectively. From these, the comparable anisotropy ratio and frictional ratio show the degree of elongation when compared to that of a globular protein.

The combination of analysing scattering and AUC improves the reliability of the data relative to that obtained by the use of either technique on its own. However, the additional structural parameters obtained by scattering make it the more powerful tool. AUC can then be employed to provide confirmation of the structural insight determined by scattering, and to check for sample polydispersity (Perkins *et al.*, 2005).

Scattering and AUC data interpretation is much improved by the use of molecular modelling based on available atomic structures (Chapter 3). This approach is particular useful in the study of antibodies, where the constituent Fab and Fc fragments have been crystallised, but crystallisation of the intact antibody is not possible (Perkins & Bonner, 2008).

Chapter Three

Constrained Biomolecular Modelling

3.1 Introduction to protein structure

3.1.1 Amino acids

Proteins are composed of amino acids that are linked by peptide bonds to form a polypeptide chain. It is the side chain, or R group, attached to the C_α atom that gives each residue characteristic properties and distinguishes one amino acid from the other (Figure 3.1). Amino acids are joined by the formation of a peptide bond, where the carboxyl group of one amino acid condenses with the amino group of another to eliminate water. This means that within a polypeptide chain, the first amino acid will have a free amino group and the last amino acid will have a free carboxy group, thus denoting the N- and C-termini, respectively.

The diversity of the amino acid side chains allows proteins to exhibit a great variety of structures, properties and functions. Residues can be grouped into classes by the chemical properties of their side chains (Figure 3.1). Alanine, valine, phenylalanine, methionine, proline, leucine and isoleucine all contain straight, branched or cyclic arrangements of carbon atoms and are hydrophobic amino acids. As these side chains become more extended, the amino acids become more hydrophobic. The side chain of proline is unique as it forms a bond with the N atom of the proline amino group making a cyclic structure. Within a protein, proline can restrict the flexibility of the polypeptide chain. The side chains of serine, threonine, tyrosine, asparagine, cysteine, tryptophan and glutamine are polar amino acids. Cysteine contains a free thiol goup, which can be oxidised to form covalent, disulphide bonds within protein structures. Phenylalanine, tyrosine and tryptophan have aromatic side chains. Phenylalanine is strongly hydrophobic, yet the polar groups in the side chains of tyrosine and tryptophan reduce their hydrophobic nature. Aspartic acid and glutamic acid side chains carry negative charges at pH 7. Lysine, arginine and histidine carry basic (positively charged) groups in their side chains. Histidine is the least basic of the three and since it can exchange protons near physiological pH, it often plays a role in enzymatic catalysis involving proton transfer. Glycine has a single hydrogen atom as its side chain, therefore reducing steric collision with other side chains and increasing the flexibility of the polypeptide chain (Figure 3.1).

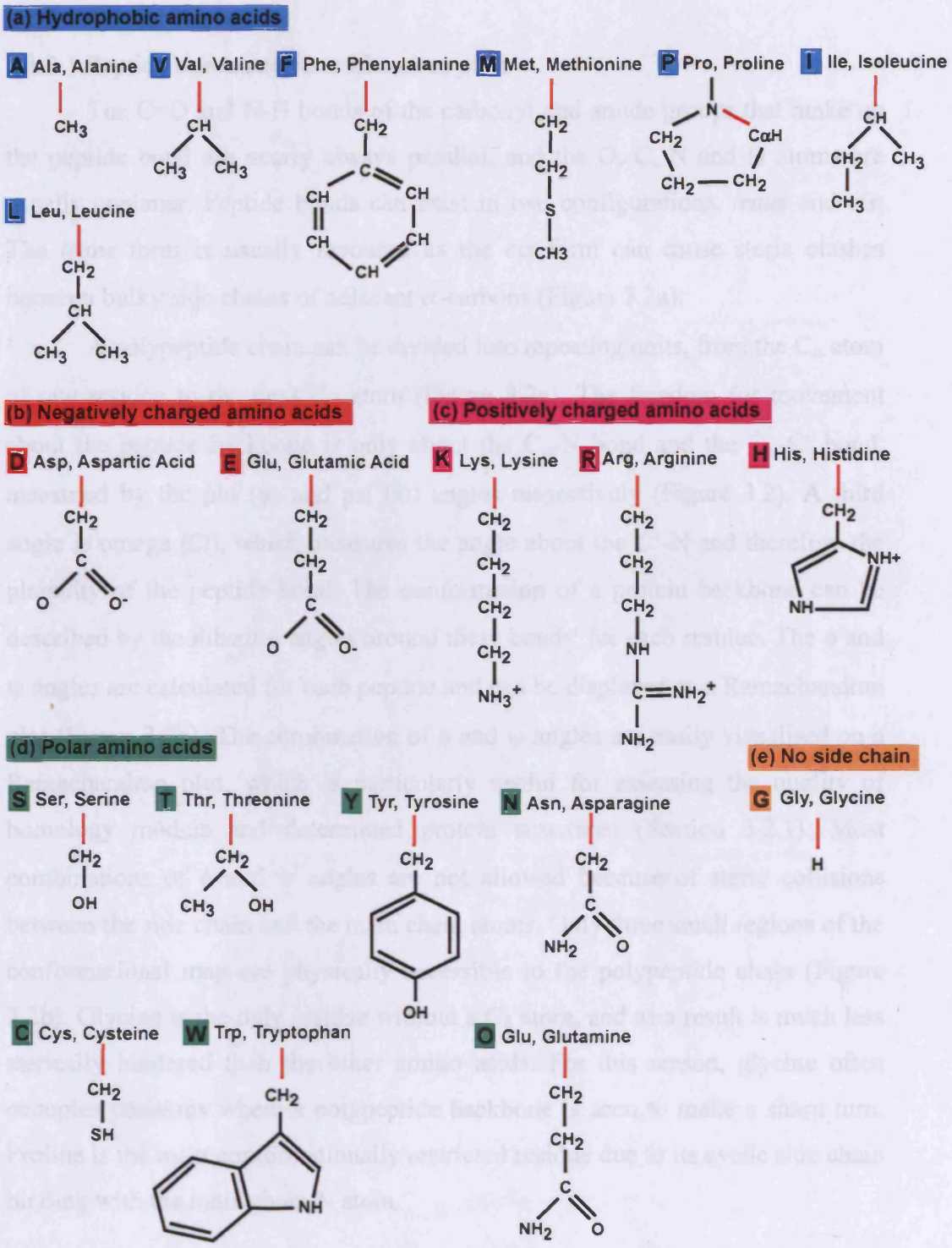

Figure 3.1 The 20 amino acid side chains

The 20 amino acids are divided into four groups (hydrophobic, polar, negatively charged and positively charged) based on the chemical properties of their side chains (R groups).

3.1.2 Secondary structure

3.1.2.1 Peptide bonds and Ramachandran plots

The C=O and N-H bonds of the carboxyl and amide groups that make up the peptide bond are nearly always parallel, and the O, C, N and H atoms are usually coplanar. Peptide bonds can exist in two configurations, *trans* and *cis*. The *trans* form is usually favoured as the *cis* form can cause steric clashes between bulky side chains of adjacent α-carbons (Figure 3.2a).

A polypeptide chain can be divided into repeating units, from the C_α atom of one residue to the next C_α atom (Figure 3.2a). The freedom for movement about the peptide backbone is only about the C_α-N bond and the C_α-C' bond, measured by the phi (ϕ) and psi (ψ) angles respectively (Figure 3.2). A third angle is omega (Ω), which measures the angle about the C'-N and therefore the planarity of the peptide bond. The conformation of a protein backbone can be described by the dihedral angles around these bonds' for each residue. The ϕ and ψ angles are calculated for each peptide and can be displayed as a Ramachandran plot (Figure 3.2b). The combination of ϕ and ψ angles are easily visualised on a Ramachandran plot, which is particularly useful for assessing the quality of homology models and determined protein structures (Section 3.2.1). Most combinations of ϕ and ψ angles are not allowed because of steric collisions between the side chain and the main chain atoms. Only three small regions of the conformational map are physically accessible to the polypeptide chain (Figure 3.2b). Glycine is the only residue without a C_β atom, and as a result is much less sterically hindered than the other amino acids. For this reason, glycine often occupies positions where a polypeptide backbone is seen to make a sharp turn. Proline is the most conformationally restricted residue due to its cyclic side chain binding with the main chain N atom.

3.1.2.2 Alpha helices and Beta sheets

α-helices and β-sheets form the secondary structure of proteins. The α-helix is defined by a stretch of residues with ϕ and ψ angle pairs of approximately -60° and -50°, respectively. This corresponds to the allowed region in the bottom left quadrant of the Ramachandran plot (Figure 3.2b). In contrast to

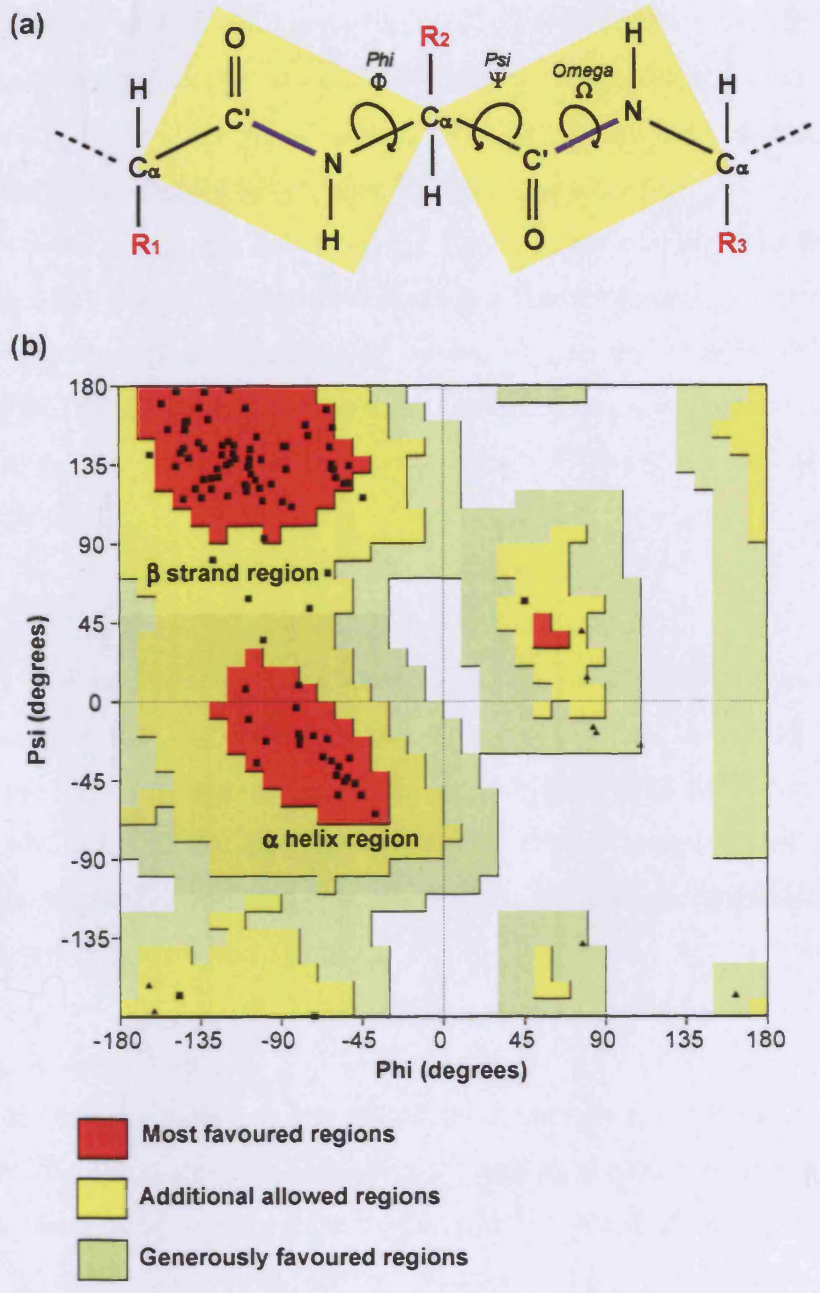

Figure 3.2 Peptide bond angles

(a) Bond dihedral angles are shown in a peptide with R_1, R_2 and R_3 sidechains. The peptides are shown in the trans conformation. The NH-Cα is defined by the phi (ϕ) angle. Cα-CO is defined by the psi (ψ) angle. The peptide bond (purple), defined by the omega (Ω) angle, is rigid and maintained in a planar structure (yellow). (b) A Ramachandran plot is shown. A good quality model would be expected to have at least 90% of its residues in the most favoured regions. This Ramachandran plot was an output from PROCHECK, analysing the bond angles of the crystal structure of SC D1 (PDB code: 1XED; Hamburger *et al.*, 2004).

the α-helix that is built from a continuous region of the polypeptide chain, the β-sheet is formed from numerous stretches of five to ten residues that are in different sections of the polypeptide chain. β-sheet motifs have ϕ and ψ backbone angles within the upper left region of the Ramachandran plot (Figure 3.2b). Within protein structures, α-helices and β-strands are combined to form stable hydrophobic pockets within the molecule. These secondary structure elements are connected by loop regions of various lengths and conformations. These loop regions almost always occur on the surface of the molecule. It is the loop regions of Ig V-type domains that form antibody CDRs (Section 1.2.2.2; Figure 1.5; Figure 1.13)

3.1.2.3 Domain structure

Within globular protein structures, several secondary-structure motifs combine to form compact regions of structure known as domains. A domain can be defined as part of a polypeptide chain that can be folded into a stable tertiary structure independently of the rest of the protein. Protein domains can also evolve independently to the rest of the protein. Protein domains can range in size, from around 25 to 250 amino acid residues.

3.2 Biomolecular modelling

Biomolecular modelling is the creation of models that represent the structure of biological molecules. An atomic coordinate model is the most detailed representation of a protein molecule, and can be obtained by NMR (Section 2.1.2) or X-ray crystallography (Section 2.1.3). The models are deposited in the Protein Data Bank (PDB) with X, Y and Z coordinates specifying the position of every atom. The PDB currently contains over 49,000 structures (February 2008; http://www.pdb.org) of which approximately 42,000 are X-ray crystal structures. If the three-dimensional structure of a protein of interest (the target) is not available in the PDB, structural information can often be obtained by analysis of a homologous protein with known structure (the template). This is a useful substitute if high-resolution structural data cannot be obtained, especially when combined with low-resolution experimental structural data such as SAS and AUC (Chapter 2).

3.2.1 Homology modelling

Homology modelling is a technique that uses known sequences and structures of homologous domains as a template to create a three-dimensional structure of the target domain.

The accuracy of homology models of protein domains generally depends on two main criteria. (i) Finding the best template structure with highest sequence similarity and similar secondary structure. The most accurate models are obtained with the fewest residues that need to be inserted into or deleted from the template in order to build the model. Template sequences and structures can be found by searching the PDB using web-based search tools such as BLAST, SwissProt, Pfam and Interpro. (ii) Generating a correct alignment of the target and template sequences. The optimal alignment for homology modelling should be identical to the alignment achieved if both structures were known and structurally aligned. Automatic sequence alignments can be obtained using methods such as CLUSTALW. However, they do not always achieve the optimal sequence alignment needed for homology modelling since they are concerned with sequence and not structure, and manual inspection and editing is necessary.

The secondary structure of the template structure is determined using the programme DSSP (Dictionary of Secondary Structure of Proteins). It allows secondary structure elements and loops within domains to be defined by analysing the pattern of hydrogen bonds within the structure and the Cα-Cα distances, bond dihedral angles and solvent exposure for each residue. Each residue is then classified into the secondary structure element it comprises; turns (T), helices (H, G or I), β-bridges (B), extended β-stands (E), main-chain bend (S) or unclassified random coil (C) (Kabsch & Sander, 1983). In homology modelling, residue insertions and deletions should be restricted to structurally variable regions such as loop regions, and should be omitted from the structurally conserved regions and secondary structure elements defined by DSSP.

Homology domain models can be generated using modelling software such as MODELLER (Sali & Blundell, 1993). MODELLER creates homology models by assessing the conserved and variable regions of the template structure and then using molecular dynamics to fit the target sequence to it, based on the

sequence alignment. Restraints are imposed to weight the structurally conserved regions (i.e. secondary structure elements) more highly than those of high variability (i.e. loop regions).

3.2.1.1 Limitations of homology modelling and assessing the structure

There are several things that can affect the accuracy of a homology model. It is important to realise that any NMR or X-ray experimental structure that is used as a template is itself a model based on experimental data. Therefore, a model of a structure obtained from homology modelling is in fact a model based on another model. The least accurate regions of the model will be the insertions and deletions within structurally variable regions as they have not been built directly from the template structure.

Homology models can be refined by energy minimisation, which involves moving atoms within a structure in order to find the minimum energy conformation. MODELLER automatically assesses and refines the domain models by energy minimisation, and therefore attempts to remove bad contacts and other structural artefacts such as those arising from changes in residue side-chain size, peptide backbone strain or non-optimal loop conformations.

PROCHECK can then be employed to check the validity of the model (Laskowski *et al.*, 1993). PROCHECK assesses the PDB atomic coordinate file using stereochemical parameters to evaluate bond lengths and bond angles within every residue of the protein. The output of PROCHECK includes a Ramachandran plot (Figure 3.2b).

3.3 Constrained scattering modelling

Modelling extends the scattering and AUC analyses by determining the three-dimensional structure that accounts for the observed scattering curve and calculated experimental parameters. Individual homology models of each domain are created (Section 3.2.1), and then combined to produce around ten thousand models that encompass all possible orientations of the intact protein. The X-ray and neutron scattering curves are then calculated for each model, and their sedimentation coefficients are determined for comparison with that experimentally obtained. However, unique structures for the protein are not possible due to the random orientation of the proteins in solution and the radial

averaging of the scattering curve. Consequently, constrained scattering modelling is particularly efficient at ruling out poor-fit structures and determining a family of best-fit structures that are compatible with the experimental data. By constraining the 10,000 models by filtering out those that do not agree with the experimental data, the best-fit models can be determined that represent the structure at medium-resolution of the protein. The determined best-fit solution structure models are based on experimental data and therefore qualify for PDB deposition, and they can provide biologically important information on the proteins (Perkins *et al.*, 2008).

3.3.1 Analysis of glycoprotein composition

Prior to solution scattering modelling, essential information is determined from the amino acid and monosaccharide composition of the protein using the program SLUV (Perkins, 1986). The partial specific volume $\bar{v}$ and molecular mass are determined. For concentration measurements, the extinction coefficient (at a wavelength of 280 nm) is calculated from the tryptophan, tyrosine and cysteine content of the protein (Perkins, 1986).

3.3.2 Creating atomic models

If atomic resolution structures are available for the domains of the protein being studied, these are used as the basis for the models created by the automated modelling procedures. If there are no atomic resolution structures, the individual domains are modelled using the known structures of homologous domains as templates (Section 3.2.1).

In order to create an arbitrary series of models that test all possible orientations of the protein and domain arrangement, each peptide outside the fixed domain structure is designated as a linker, and a library of different conformations is generated for each linker. A library of around 5,000 random linker conformations for each linker peptide are generated using INSIGHT II 98.0 molecular graphics software with BIOPOLYMER, DISCOVER and HOMOLOGY modules (Accelrys, San Diego, CA, USA). A wide range of conformations are created by subjecting the peptide structure to molecular dynamics for 300 iterations at a temperature of 773K. After a temperature equilibration step of 5,000 fs, the simulation is run for 500,000 fs. The linker

structure is then saved every 100 fs to produce the linker library (Section 4.4.5). In the automated modelling procedure, linkers are then randomly selected from the linker library to join the domains together to generate a series of complete models (Figure 3.3a). In a typical randomised modelling search, 10,000 trial-and-error models are generated.

If the protein being studied contains putative glycosylation sites, a three-dimensional structure for the N- and O-linked oligosaccharide chains is positioned at the glycosylation sites on the protein. When present, the modelling of antibody carbohydrate chains presented in this thesis generally contained N-linked biantennary complex-type carbohydrates with a Man3GlcNAc2 core and two NeuNAc.Gal.GlcNAc antennae, and O-linked NeuNAc.Gal.GalNAc oligosaccharides.

3.3.3 Debye sphere modelling

To calculate a theoretical scattering curve from an atomic model of a protein, the protein is first converted into a sphere model consisting of small non-overlapping spheres of the same total volume as the original atomic model (Figure 3.3b). To do this, the atomic coordinate model is placed within a three dimensional grid of cubes. The size of the cubes and the number of atoms that each cube contains is user-defined, and both are tested empirically by trial and error, so that after the grid transformation the resulting sphere model has a volume as close to the unhydrated volume of the protein as possible. The unhydrated volume is calculated from the protein composition using SLUV (Section 3.3.1). The length of the cube side that is used to produce an unhydrated sphere model is typically between 0.5 and 0.6 nm, which is much less than the normal resolution of a normal scattering curve of around 3 nm (Section 2.2.2). Once a grid size has been determined, it is kept constant for the automated curve fit searches.

Neutron scattering experiments do not detect a hydration shell around proteins, so unhydrated sphere models are used for neutron scattering curve calculations (Figure 3.3b). In contrast, the hydration shell is detected in X-ray scattering experiments, and therefore the calculation of an X-ray scattering curve requires that the unhydrated sphere model is modified to consider the hydration

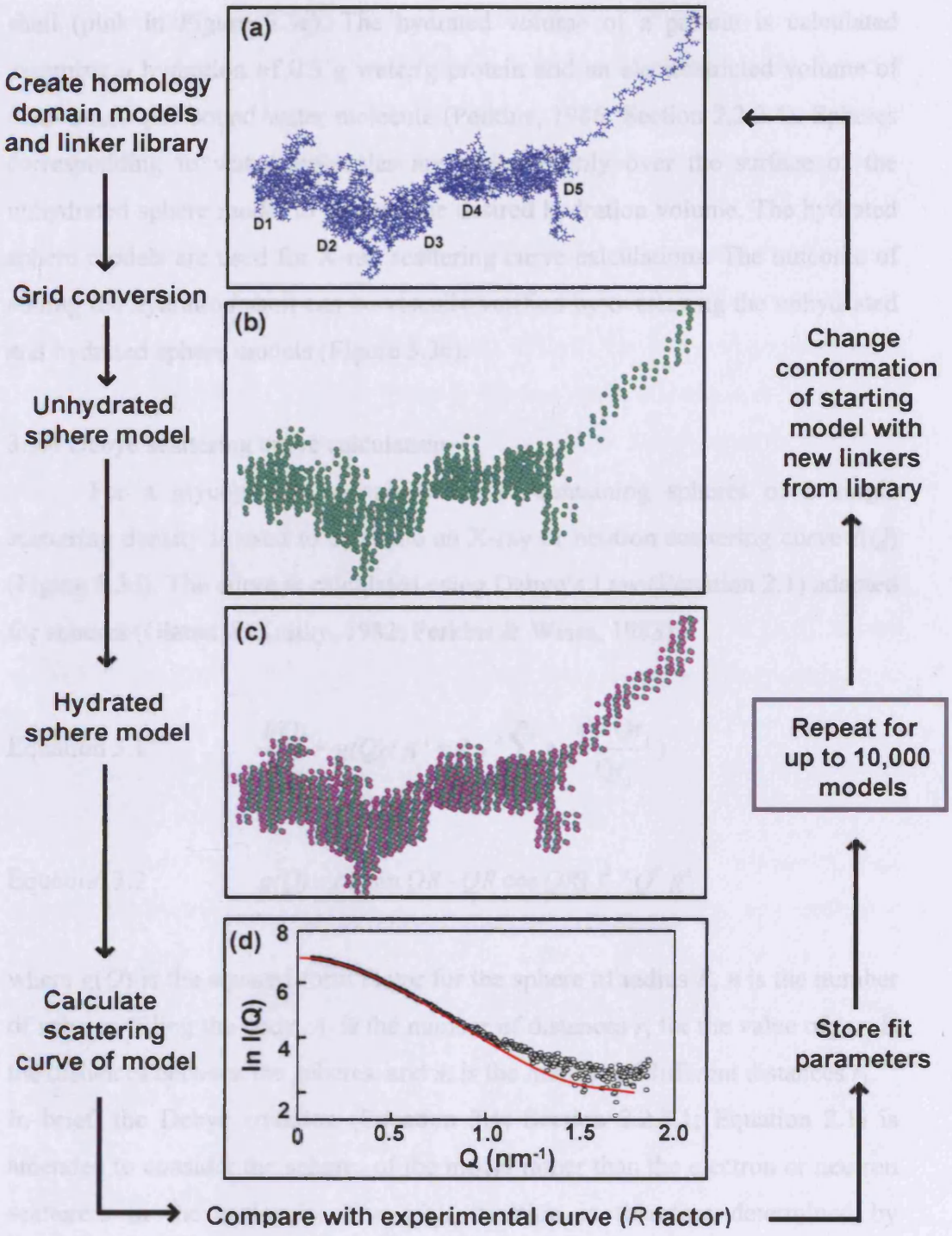

Figure 3.3 Overview of the constrained modelling procedure

(a) A model for SC is shown (blue), with D1-D5 in an extended arrangement. (b,c) The unhydrated (green) and hydrated (pink) sphere models are shown. (d) The calculated scattering curve of the model (red) is compared to the experimental scattering curve (open circles) to determine the goodness-of-fit *R* factor.

shell (pink in Figure 3.3c). The hydrated volume of a protein is calculated assuming a hydration of 0.3 g water/g protein and an electrostricted volume of 0.0245 nm^3 per bound water molecule (Perkins, 1986; Section 2.2.2.4). Spheres corresponding to water molecules are added evenly over the surface of the unhydrated sphere model to achieve the desired hydration volume. The hydrated sphere models are used for X-ray scattering curve calculations. The outcome of adding the hydration shell can be visually verified by overlaying the unhydrated and hydrated sphere models (Figure 3.3c).

3.3.4 Debye scattering curve calculation

For a glycoprotein, a sphere model containing spheres of a single scattering density is used to calculate an X-ray or neutron scattering curve $I(Q)$ (Figure 3.3d). The curve is calculated using Debye's Law (Equation 2.1) adapted for spheres (Glatter & Kratky, 1982; Perkins & Weiss, 1983):

Equation 3.1
$$\frac{I(Q)}{I(0)} = g(Q)\left(n^{-1} + 2n^{-2}\sum_{j=1}^{m} A_j \frac{\sin Qr_j}{Qr_j}\right)$$

Equation 3.2
$$g(Q) = (3(\sin QR - QR\cos QR))^2 / Q^6 R^6$$

where $g(Q)$ is the squared form factor for the sphere of radius R, n is the number of spheres filling the body, A_j is the number of distances r_j for the value of j, r_j is the distances between the spheres, and m is the number of different distances r_j.
In brief, the Debye equation (Equation 3,1; Section 2.2.2.1; Equation 2.1) is amended to consider the spheres of the model rather than the electron or neutron scatterers in the molecule. The $g(Q)$ function is therefore determined by calculating the distances r from each sphere to the remaining spheres and summing the results (Equation 3.2). The single density scattering curve calculation is applicable to proteins and glycoproteins with low carbohydrate content, provided that the comparisons between the experimental data sets are consistent for the neutron and X-ray curves.

If a glycoprotein has a high carbohydrate content, differences between the scattering densities of the protein and its carbohydrate moieties may give rise to

systematic deviations between the neutron and X-ray curves. The glycoprotein can then, in principle be modelled using two-density sphere models. This was not necessary for the work presented in this thesis as no major discrepancies were seen between the X-ray and neutron modelled curves for the glycoproteins studied.

Synchrotron X-ray cameras utilise a pin-hole configuration that does not produce distortion of the beam (Figure 2.3), so a calculated X-ray curve does not have to be corrected for instrumental effects in order for it to be compared with an experimental curve. Although neutron cameras such as those at ISIS also use pin-hole geometries, their dimensions are larger than X-ray cameras and they also use longer wavelengths to maximise the available neutron flux (Figure 2.4). For these reasons, it is necessary to apply instrumental corrections to neutron scattering curves of the models. For LOQ, a Gaussian function based on a 10% wavelength spread and beam divergence of 0.016 radians is applied to the modelled neutron curves during the automated modelling procedure (Ashton *et al.*, 1997).

3.3.5 Evaluation of models

Each model is evaluated using several criteria. (i) The volume of each sphere model, indicated by the number of spheres generated from the coordinates during the grid transformation, is used to determine whether the automated modelling procedure has produced significant steric overlap of the protein domains. Significant overlap means that the number of spheres is much reduced (usually >5%) compared to that predicted from its composition. (ii) Each model is also evaluated by comparison of its calculated scattering curves to the experimental data. X-ray and neutron R_G and R_{XS} values are calculated from each modelled scattering curve over the same Q ranges used for the experimental measurements. (iii) An R factor is calculated for a quantitative comparison of each model scattering curve $I(Q)_{cal}$ against experimental scattering data $I(Q)_{exp}$ over the whole Q range, where $I(0)_{cal}$ is set as 1000:

Equation 3.3
$$R = \frac{\sum | I(Q)_{exp} - I(Q)_{cal} |}{\sum | I(Q)_{exp} |} \times 100\%$$

The R factor values reflect the goodness-of-fit between the calculated and experimental curves and are analogous to that used for the refinement of crystallographic models. R factor values should be less than 10% for a good fit. Using these three criteria, it is possible to minimise the number of models that have to be considered from a large automated search. Models that do not conform to the experimentally defined ranges of model volume, R_G and R_{XS} are rejected using filters in EXCEL, and the remaining models are further evaluated. The structure of the final models are manually visualised to ensure that they are stereochemically viable. Also, the modelled scattering curves are manually compared to the experimental curves and their sedimentation coefficients are calculated (Section 3.3.6) to ensure that there is no discrepancy between the modelled and experimental parameters. The theoretical sedimentation coefficient $s^o_{20,w}$ of the best-fit models are determined directly from the PDB coordinates by using the programme HYDROPRO (Garcia de la Torre *et al.*, 2000). Comparison of the sedimentation coefficients of the models with that experimentally obtained by SV experiments provides another means of validating the scattering modelling. The best-fit models are then determined based on their R factors. The models can be compared with the structures already obtained by EM, NMR or X-ray crystallography. The α-carbon coordinates are then deposited in the PDB and given accession codes (Perkins & Bonner, 2008).

Chapter Four

Solution Structure of Human Secretory Component: Implication for Biological Function

4.1 Introduction

Secretory component (SC) is involved in the formation of SIgA and its transportation to mucosal secretions, where SIgA mediates a major part of mucosal immune defence (Nagler-Anderson, 2001). IgA is produced locally as (generally) dimeric IgA (dIgA). Subsequent transport of dIgA (and to a lesser extent IgM) across the epithelium is ensured by the polymeric immunoglobulin receptor (pIgR) which is expressed on the basolateral surface of epithelial cells. Following cleavage at luminal surfaces, SIgA is released as a complex of dIgA bound to the cleaved extracellular portion of the pIgR which is now known as SC (Section 1.4.1; Figure 1.11).

SC consists of five immunoglobulin-like domains (D1-D5) to which up to seven glycan chains are attached (Figure 4.1) that account for 15-25% of its total molecular weight (Cleveland *et al.*, 1991). SC bound to dIgA protects the Ab against proteolytic digestion (Crottet & Corthésy, 1998), and governs anchoring of SIgA at mucosal surfaces (Phalipon *et al.*, 2002). Free SC is also found as such in secretions and is now recognized as an active antibacterial participant. Recombinant human SC produced from transfected Chinese hamster ovary (CHO) cells behaves identically to that purified from colostrum and its neutralizing properties have been shown to rely on the presence of carbohydrate moieties on SC (Perrier *et al.*, 2006).

So far, the question of the three-dimensional domain arrangement of SC has not been considered, even though this is essential to appreciate molecular mechanisms that are involved in mucosal immunity. Its high glycosylation and the presence of long inter-domain linkers suggest that SC may be too flexible to be crystallised intact (Figure 4.1). In a situation such as this, X-ray and neutron scattering and AUC in combination with constrained modelling techniques will provide a model of the domain arrangement of SC (Perkins *et al.*, 2008). By these methods, a solution structure for recombinant human SC has been determined. This is verified with data for the D1-D3 and D4-D5 fragments of SC. SC forms a compact domain arrangement in solution, and this structure rationalises many of the biological roles of both free and bound SC for the first time.

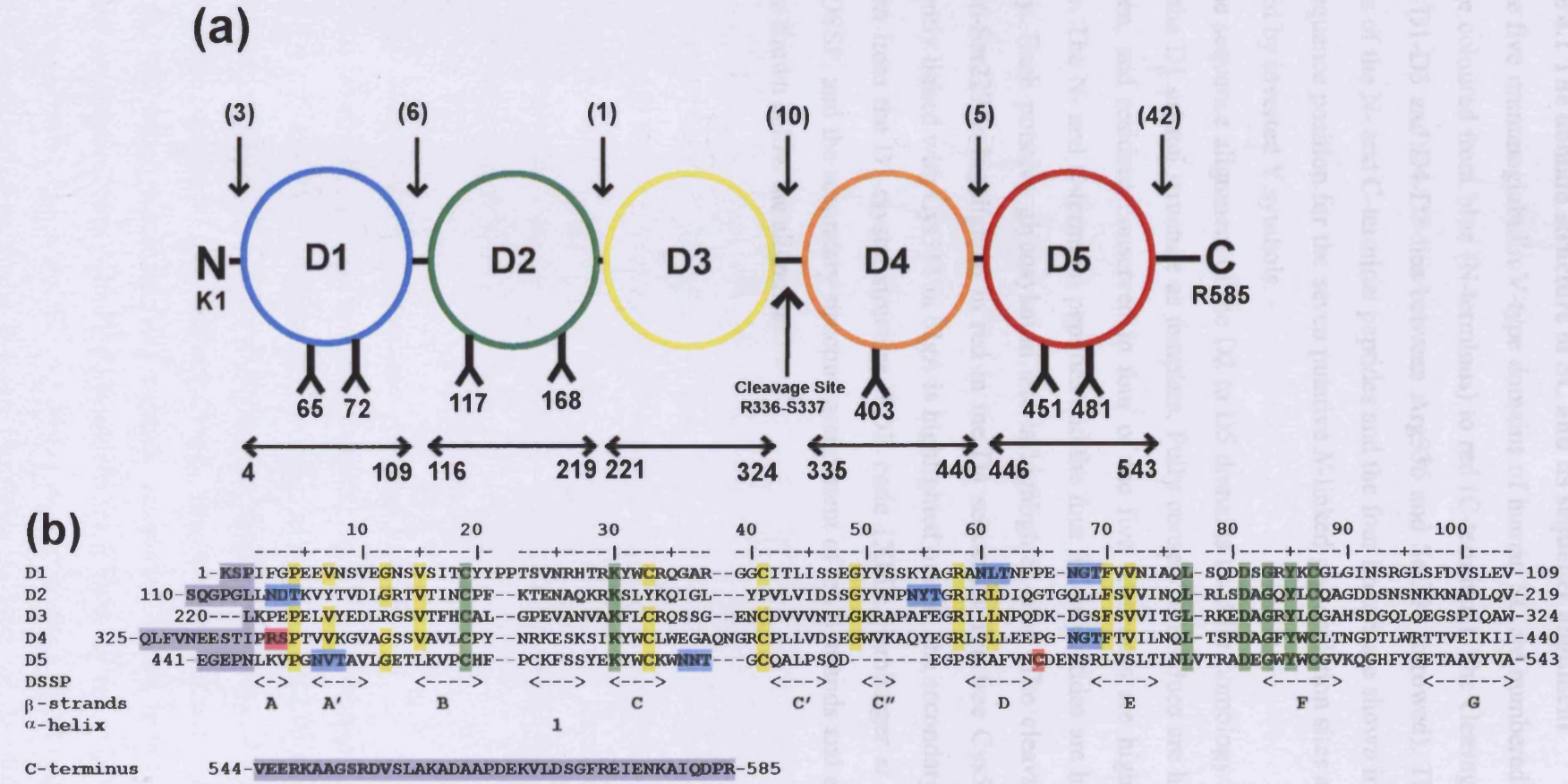

Figure 4.1 (Legend overleaf)

Figure 4.1 The domain structure of SC and its sequence alignment

(a) The five immunoglobulin V-type domains of human SC are numbered D1 to D5 and are coloured from blue (N-terminus) to red (C-terminus). The cleavage site that yields D1-D3 and D4-D5 lies between Arg336 and Ser337 (arrowed). The residue lengths of the N- and C-terminal peptides and the four linkers are shown in brackets. The sequence position for the seven putative *N*-linked glycosylation sites are shown, denoted by inverted Y symbols.

(b) The sequence alignment of the D2 to D5 domains used for homology modelling using the D1 crystal structure as template. Fully conserved residues are highlighted in green, and residues conserved in four of the five sequences are highlighted in yellow. The N- and C-terminal peptides and the four linker peptides are highlighted in grey. Each putative glycosylation site is highlighted blue. The cleavage site at Arg336-Ser227 is highlighted in red in the D4 sequence. The free Cys502 that is covalently linked with Cys311 in SIgA is highlighted in red. The secondary structure is taken from the D1 crystal structure (PDB code 1XED; Hamburger *et al.*, 2004) from DSSP, and the secondary structure assignment of ten β-strands and a single α-helix is shown below the alignment.

4.2 Results

4.2.1 Characterisation of SC and its fragments

In SC, sequence similarities between D1 and the other four SC domains (D2-D5) show that all belong to the immunoglobulin (Ig) variable (V)-type classification (Figure 4.1a; Section 1.2.2). For clarification, the domains and interdomain linkers are depicted in Figure 4.1b. As glycosylation and proper disulphide bonding are essential features of SC, it was produced in CHO cells which yields SC that is indistinguishable when compared to colostrum SC (Crottet & Corthésy, 1999; Perrier *et al.*, 2006). Purified recombinant human SC resulted in a clean single band at an apparent molecular mass of 75 kDa when analysed by silver staining and immunodetection (Figure 4.2, lanes 1, 3, 5 and 7). This molecular mass agrees well with the sequence derived value of 79.6 kDa. Upon storage for several months at 4°C, two additional bands at 45 kDa and 30 kDa appear (Figure 4.2, lanes 2, 4, 6 and 8). In Western blots, detection with polyclonal antisera specific for D1 and D5 established that the bands at 45 kDa and 30 kDa represent cleavage of full-length SC into D1-D3 and D4-D5 fragments. This was confirmed by N-terminal sequencing of fragment D4-D5 that identified the sequence SPTVVKGVAG corresponding to the beginning of D4, and indicating cleavage between Arg336 and Ser337 in the SC sequence (SWISSPROT code P01833). The PeptideCutter web tool (Gasteiger *et al.*, 2005), available from the ExPASy server suggested that the enzymatic cleavage may be caused by one of trypsin, thrombin, clostripain or Arg-C proteinase. This unexpected observation provided us with three different polypeptides that were used to generate the results presented below.

4.2.2 X-ray and neutron Guinier analyses of SC and its fragments

Solution scattering was used to investigate the domain arrangements in SC and the D1-D3 and D4-D5 fragments at concentrations between 0.40 mg/ml to 2.7 mg/ml (Section 4.4.2). The acquired X-ray data showed no change during the ten successive time frames of each data acquisition, from which it was concluded that no X-ray induced radiation damage had occurred. Accordingly the ten time frames were averaged for data analyses. SDS-PAGE analyses of the samples before and after data acquisitions showed that, while SC and D4-D5 remained intact, further cleavages in D1-D3 occurred to give single domain fragments as well as the three-domain fragment.

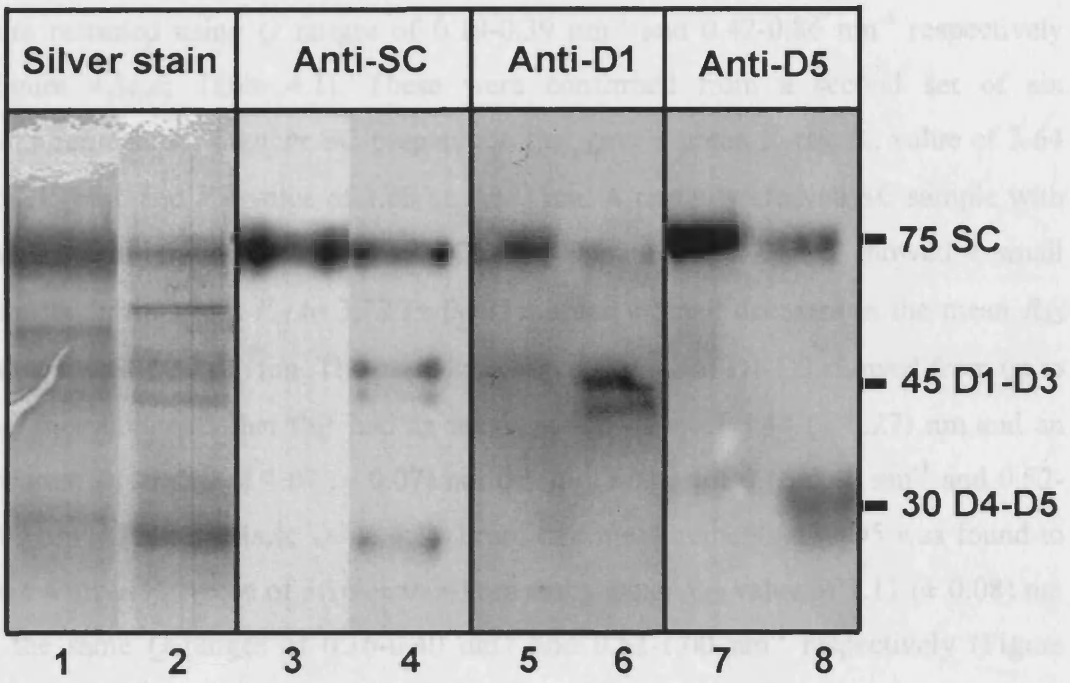

Figure 4.2 SDS-PAGE and Western blot analyses of SC and its fragments

SDS-PAGE and Western blot analyses of SC (75 kDa; lanes 1, 3, 5, 7) and cleaved SC (lanes 2, 4, 6, 8) containing a mixture of SC, D1-D3 (major band at 45 kDa) and D4-D5 (30 kDa). Lanes 1 and 2 represents a silver stain of the SDS-PAGE analysis, while the three successive panels represent Western blots of the gel in lanes 1 and 2 using (lanes 3 and 4) anti-SC, (lanes 5 and 6) anti-D1 and (lanes 7 and 8) anti-D5 antibodies.

At low Q values, Guinier analyses of ln $I(Q)$ against Q^2 resulted in linear plots for SC and its two fragments, from which the radius of gyration (R_G) values were obtained (Figure 4.3a; Figure 4.4a,b). At larger Q values, analyses of ln $I(Q)Q$ against Q^2 also resulted in linear plots, from which the R_G of the cross-section (R_{XS}) were obtained (Figure 4.3c; Figure 4.4c,d). From seven measurements of SC, a mean X-ray R_G value of 3.53 (± 0.03) nm and mean R_{XS} value of 1.76 (± 0.08) nm were recorded using Q ranges of 0.19-0.39 nm^{-1} and 0.42-0.86 nm^{-1} respectively (Figure 4.3a,c; Table 4.1). These were confirmed from a second set of six measurements for another SC preparation that gave a mean X-ray R_G value of 3.64 (± 0.17) nm and R_{XS} value of 1.63 (± 0.12) nm. A partially cleaved SC sample with an estimated content of 33% SC, 33% D1-D3 and 33% D4-D5 showed a small increase in the mean R_G to 3.77 (± 0.17) nm and a small decrease in the mean R_{XS} value to 1.60 (± 0.61) nm. The partially cleaved sample of D1-D3 showed from up to five measurements that this had an apparent R_G value of 3.44 (± 0.27) nm and an apparent R_{XS} value of 1.07 (± 0.07) nm using Q ranges of 0.16-0.40 nm^{-1} and 0.52-1.00 nm^{-1} (Figure 4.4a,c; Table 4.1). From five measurements, D4-D5 was found to have a mean R_G value of 3.16 (± 0.04) nm and a mean R_{XS} value of 1.11 (± 0.08) nm in the same Q ranges of 0.16-0.40 nm^{-1} and 0.52-1.00 nm^{-1} respectively (Figure 4.4b,d; Table 4.1). The Guinier analyses were confirmed by the ratio of the molecular mass values of SC, D1-D3 and D4-D5 using their Guinier $I(0)/c$ values, where c is the protein concentration (Perkins *et al.*, 1998). The observed $I(0)/c$ values were 0.0077, 0.0023 and 0.0041 for SC, D1-D3 and D4-D5 respectively, relative to the Lupolen standard, and these were in the ratio of 1 : 0.30 : 0.53. The sequence-derived molecular weight ratios for SC, D1-D3 and D4-D5 are 1 : 0.57 : 0.43 in that order. The value for D1-D3 is notably lower than expected from the sequence, which is attributed to the presence of single-domain cleavage products, while that for D4-D5 is in good agreement with that expected from the sequence.

Interestingly SC appeared more compact than either D1-D3 or D4-D5. Shape information on SC, D1-D3 and D4-D5 was obtained from the anisotropy ratio, R_G/R_O (where R_O is the R_G value of a sphere of equal volume to that of the hydrated glycoprotein). The ratios, calculated from the X-ray R_G values, were 1.44-1.49, 1.69 and 1.70 for SC, D1-D3 and D4-D5 in that order (Table 4.1). As the anisotropy ratio for typical globular protein is close to 1.28 (Perkins, 1988), this shows that all three

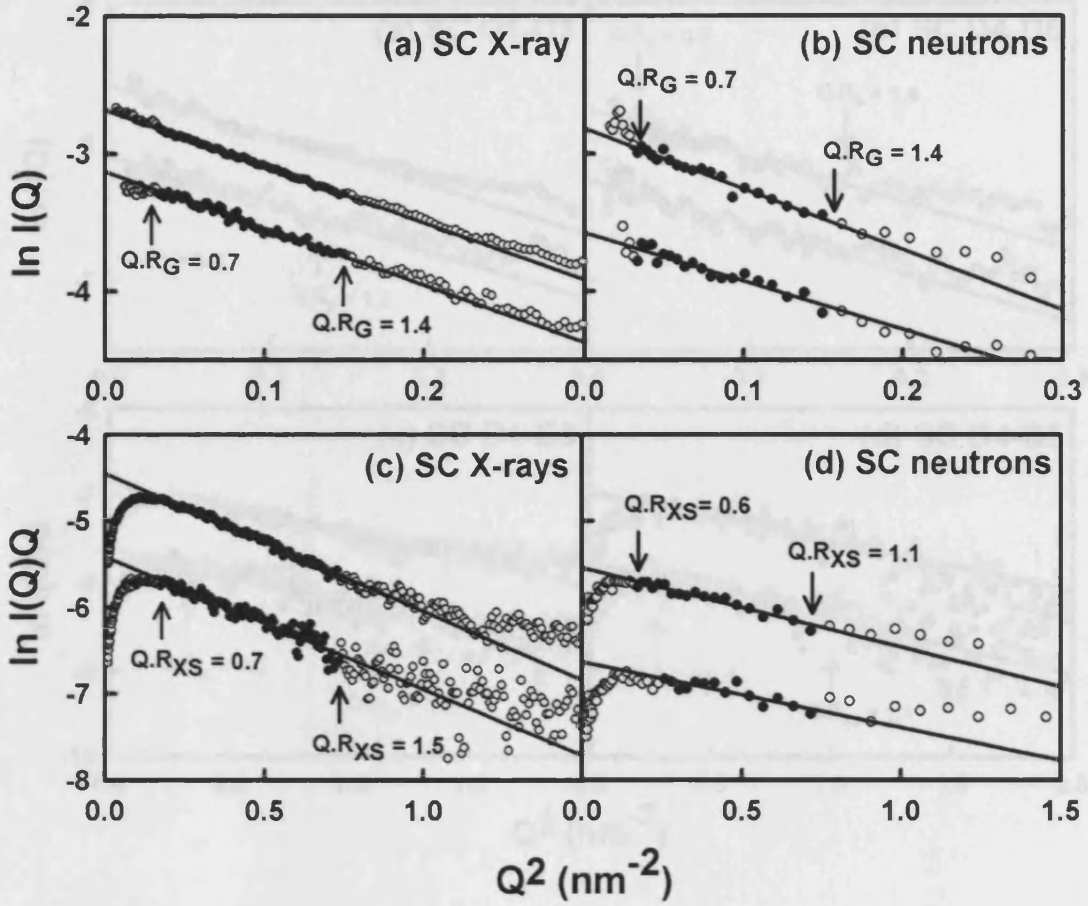

Figure 4.3 Guinier R_G and R_{XS} analyses for SC

Data points included in the Guinier fits are indicated by filled circles between the arrowed QR_G and QR_{XS} ranges. The Q ranges for SC are 0.18-0.39 nm^{-1} (R_G) and 0.42-0.86 nm^{-1} (R_{XS}). (a, c) X-ray fits for SC at a concentration of 1.3 mg/ml. (b, d) Neutron fits for SC at concentrations of 2.0 mg/ml (upper) and 1.3 mg/ml (lower). In panels (a, c, d) the two Guinier plots are arbitrarily displaced on their vertical axes for reason of clarity.

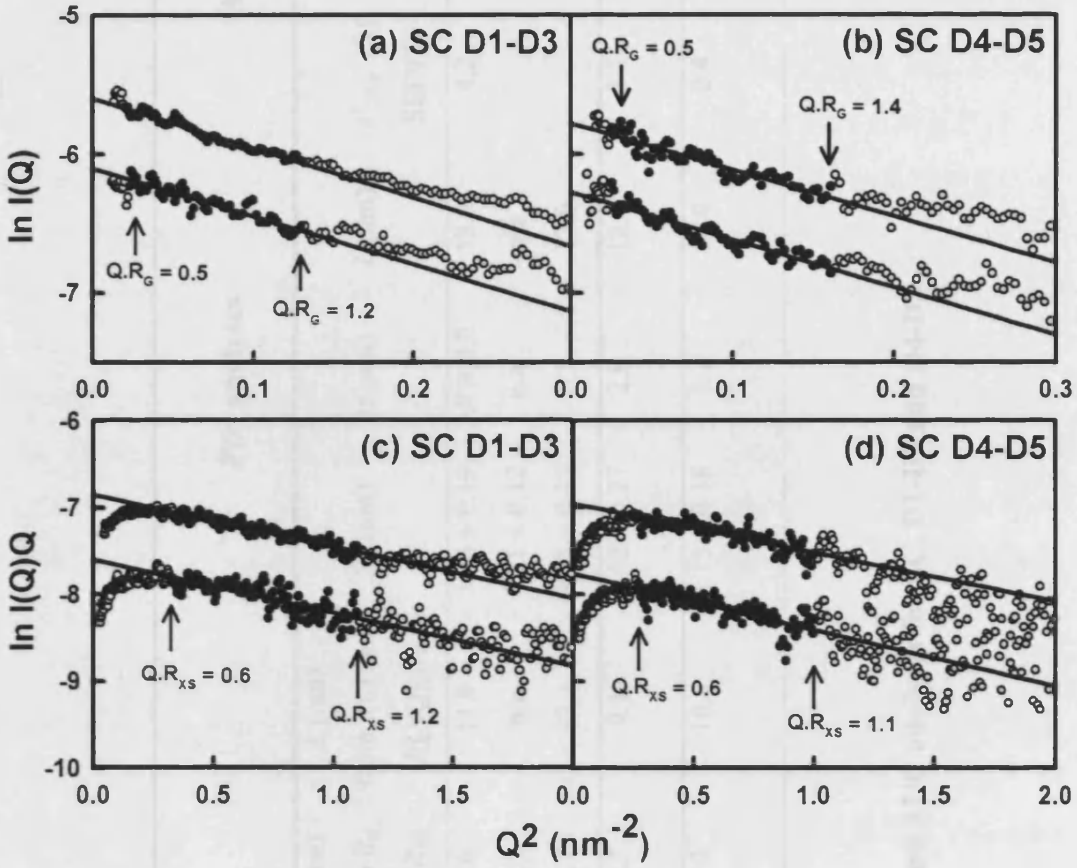

Figure 4.4 Guinier R_G and R_{XS} analyses for D1-D3 and D4-D5

Data points included in the Guinier fits are indicated by filled circles between the arrowed QR_G and QR_{XS} ranges. The Q ranges for D1-D3 are 0.15-0.36 nm^{-1} (R_G) and 0.58-1.07 nm^{-1} (R_{XS}), and those for D4-D5 are 0.16-0.40 nm^{-1} (R_G) and 0.52-1.00 nm^{-1} (R_{XS}). (a, c) X-ray fits for D1-D3 at concentrations of 1.6 mg/ml (upper) and 1.2 mg/ml (lower). (b, d) X-ray fits for D4-D5 at concentrations of 0.7 mg/ml (upper) and 0.6 mg/ml (lower). In panels (b, c, d), the two Guinier plots are arbitrarily displaced on their vertical axes for reason of clarity.

Glycoprotein	Guinier analyses					*P(r)* analyses			AUC analyses		
	R_G (nm)	R_{XS} (nm)	R_G/R_0	L (nm) from R_G and R_{XS}	L (nm) from $I(0)$ and $[I(Q)Q]_{Q\to 0}$	R_G (nm)	M (nm)	L (nm)	$s^o_{20,w}$ (S) SEDFIT	$s^o_{20,w}$ (S) (DCDT+)	f/f_o
SC (X-ray)	3.53 ± 0.03	1.76 ± 0.08	1.44	10.6	11.9	3.66 ± 0.19	3.8 ± 0.2	13.0	4.2	4.0 (int)	1.55
SC (X-ray)	3.64 ± 0.17	1.63 ± 0.12	1.49	11.3	n.a	3.71 ± 0.12	n.a.	n.a.		3.9 (abs)	
SC (Neutron)	3.63 ± 0.28	1.30 ± 0.10	n.a.	11.7	13.1	3.73 ± 0.17	3.2	12.0			
D1-D3 (X-ray)	3.44 ± 0.27 †	1.07 ± 0.07 †	1.69	11.3	9.1	3.52 ± 0.27	2.9	12.0	3.1	n.a.	1.42
D4-D5 (X-ray)	3.16 ± 0.04	1.11 ± 0.08	1.70	10.2	10.8	3.15 ± 0.18	2.7	10.0	2.4	2.4 (int) 2.3 (abs)	1.48

Table 4.1 Experimental X-ray and neutron scattering and AUC analyses for SC, D1-D3 and D4-D5

n.a. - data not available

† Apparent (see Section 4.2.2)

protein are elongated, yet SC is more compact than its two fragments D1-D3 or D4-D5. This suggests that the SC domains do not form an elongated arrangement as depicted schematically in Figure 4.1a.

Assuming that the protein can be represented as a cylindrical shape, the lengths of SC, D1-D3 and D4-D5 were calculated from their R_G and R_{XS} values to be 10.6-11.3 nm, 11.3 nm and 10.2 nm respectively (Section 4.4). From the *I(0)* and $[I(Q)Q]_{Q\to 0}$ values, the lengths were calculated to be 11.9 nm, 9.1 nm and 10.8 nm respectively (Table 4.1). Despite the presence of cleaved products in the D1-D3 sample, the similarity of these three lengths is again explained by postulating that SC is more compact and folded back upon itself to position the D1-D3 domains adjacent to those in D4-D5. If the end-to-end length of a single Ig fold in SC is taken to be 3.7 nm from the crystal structure of the D1 domain of SC (PDB code 1XED; Hamburger *et al.*, 2004; Figure 1.13), then SC, D1-D3 and D4-D5 would be predicted to be of lengths 18.5 nm, 11.1 nm and 7.4 nm if fully extended (although this estimate does not include the inter-domain and N- and C- terminal linkers). The disparity between the observed length of SC of 10.6-11.9 nm (Table 4.1) and the prediction of 18.5 nm shows that SC must be folded back. That the observed length of 10.2-10.8 nm for D4-D5 (Table 4.1) is longer than the predicted value of 7.4 nm may result from the additional 42 C-terminal residues in SC (Figure 4.1). The two observed lengths of 11.3 nm and 9.1 nm for D1-D3 are in good agreement with the predicted extended length of 11.1 nm.

Neutron scattering data for SC confirm the X-ray findings. Neutron data were collected for SC concentrations between 0.68 mg/ml to 1.4 mg/ml (Section 4.4.2) as a control for possible radiation damage effects on the X-ray R_G measurements, as these effects are absent in neutron scattering. The neutron data are complementary to X-ray data in that the hydration shell surrounding SC is not detectable (Perkins, 2001; Section 2.2.2.4). The high negative protein-solvent scattering contrast obtained using 100% ^2H_2O buffers acts as a control of any inhomogeneity effects caused by the 19% mass of carbohydrate in SC which has a higher scattering density than that of protein. Using the same Guinier Q ranges as those used for the X-ray data, the mean of three measurements for the R_G and R_{XS} values of SC were found to be 3.63 (± 0.28) nm and 1.30 (± 0.10) nm respectively (Figure 4.3b,d; Table 4.1). The decrease in the neutron R_{XS} value compared to the X-ray R_{XS} value is attributed to the

absence of hydration effects on the cross-sectional properties of SC (Boehm & Perkins, 2000). The length of SC was calculated from its neutron R_G and R_{XS} values to be 11.7 nm (in good agreement with the X-ray value of 10.6-11.3 nm), and from its $I(0)$ and $[I(Q)Q]_{Q\to 0}$ values to be 13.1 nm (consistent with the X-ray determined value of 11.9 nm) (Table 4.1). It was concluded from the consistency of the X-ray and neutron data that satisfactory scattering data had been acquired.

The neutron data also permitted the determination of the molecular weight of SC using its linear correlation with the $I(0)/c$ value from the Guinier R_G analysis, where c is the protein concentration. For proteins measured in ^2H_2O buffers, and normalised against a standard deuterated polymer standard, the observed $I(0)/c$ value of 0.087 (± 0.009) resulted in a molecular weight of 78 (± 8) kDa, where $M = I(0)/c \times 9 \times 10^5$ (SC labelled in Figure 2.5; Boehm *et al.*, 1999). This value is within 2 % of the sequence-derived molecular weight of 79.6 kDa, and showed that SC is monomeric in solution.

4.2.3 X-ray and neutron distance distribution analyses of SC and its fragments

The transformation of the $I(Q)$ scattering curve into the distance distribution function $P(r)$ confirmed that SC has a folded-back structure in solution. The $P(r)$ curves for SC (X-rays and neutrons) and for D1-D3 and D4-D5 (X-rays) resulted in reproducible $P(r)$ curves with single maxima (Figure 4.5; Figure 4.6). As a control of the $P(r)$ analyses, the R_G values calculated from these were compared with those determined in the Guinier analyses. These mean R_G values were 3.66 (± 0.19) nm and 3.71 (± 0.12) nm (two SC X-ray samples), 3.73 (± 0.17) nm (SC neutrons), 3.52 (± 0.27) nm (D1-D3) and 3.15 (± 0.18) nm (D4-D5), all of which were consistent with the Guinier analyses (Table 4.1). The maximum M corresponds to the most frequently occurring interatomic distance. These were 3.8 (± 0.2) nm and 3.2 nm respectively for the X-ray and neutron $P(r)$ curves for SC, and were slightly smaller at 2.9 nm and 2.7 nm for D1-D3 and D4-D5 respectively (Table 4.1). The length L is determined from the point where $P(r)$ becomes zero at large r values, and is not dependent on shape assumptions. For SC, L was determined to be 13 (± 1) nm and 12 (± 1) nm from the X-ray and neutron analyses respectively (Figure 4.5a,b; Table 4.1). For the D1-D3 and D4-D5 fragments, L was determined to be 12 (± 1) nm and 10 (± 1) nm respectively (Figure 4.6; Table 4.1). While the D1-D3 Guinier analysis

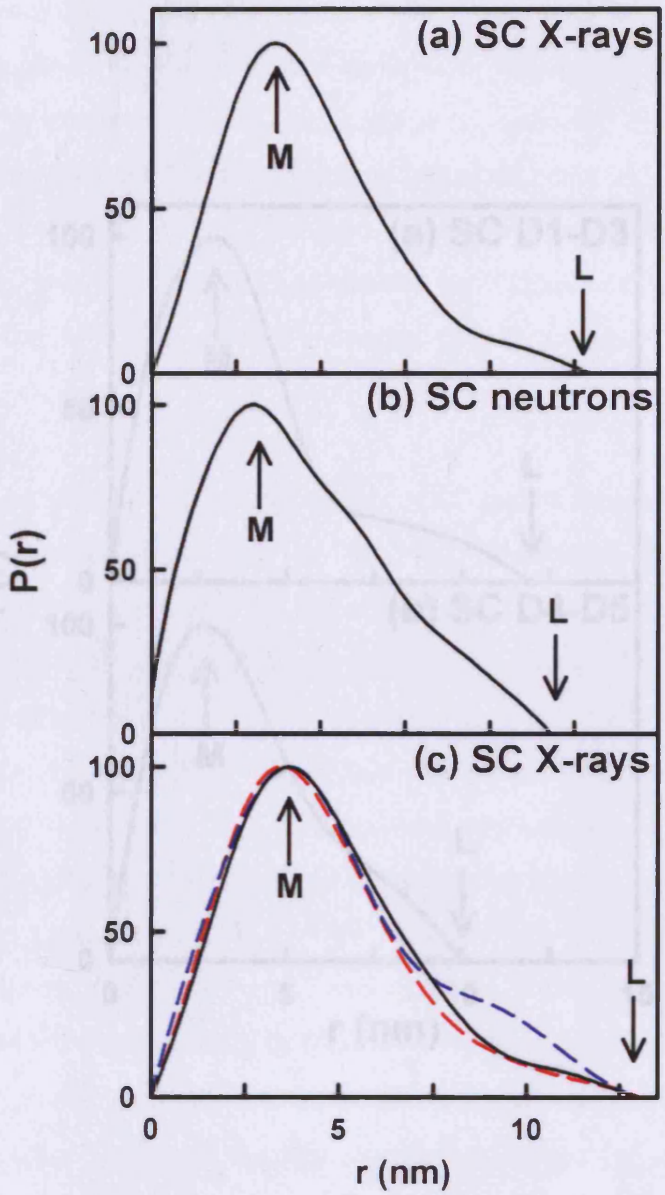

Figure 4.5 X-ray and neutron distance distribution functions *P(r)* for SC

The maximum of the *P(r)* curve at *M* depicts the most frequently occurring distance within the structure. The maximum dimension is denoted by *L* at the *r* value where the *P(r)* curve reaches 0. (a, b) The X-ray and neutron *P(r)* curves for SC resulted in *M* values of 3.6 nm and 3.2 nm respectively, and *L* values of 13 nm and 12 nm respectively. (c) *P(r)* curves are shown for three different sample of SC that are estimated from SDS-PAGE to have 100% intact SC (black continuous line, as shown in part a), 90% intact SC (red dashed line) or 66% intact SC (blue dashed line). In the cleaved samples (blue and red line), SC has been cleaved to yield a mixture of SC, D1-D3 and D4-D5. All three *P(r)* curve show the same *M* values of 3.6 nm and *L* values of 13 nm.

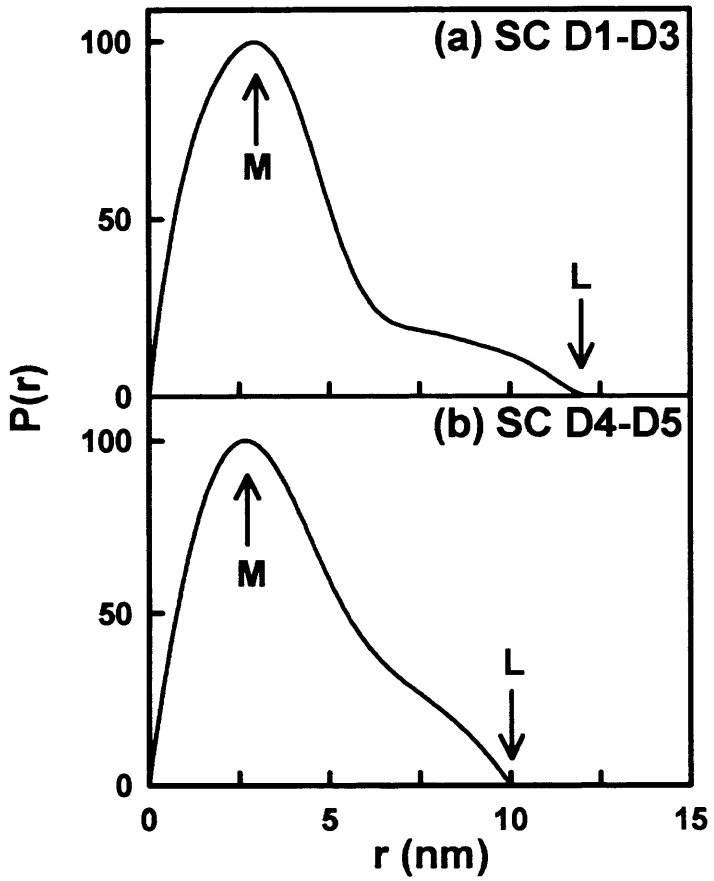

Figure 4.6 X-ray and neutron distance distribution functions *P(r)* for D1-D3 and D4-D5

The maximum of the *P(r)* curve at *M* depicts the most frequently occurring distance within the structure. The maximum dimension is denoted by *L* at the *r* value where the *P(r)* curve reaches 0. (a, b) The X-ray *P(r)* curves for D1-D3, and D4-D5 resulted in *M* values of 2.9 nm and 2.7 nm respectively, and *L* values of 12 nm and 10 nm respectively.

may be affected by cleaved products, this *P(r)* analysis is expected to reveal the maximum length of the intact protein as any cleavage products will be shorter in length. This is confirmed by the comparison of three SC *P(r)* curves for three different SC samples, each with a different degree of intact SC cleavage, estimated by SDS-PAGE (Figure 4.5c). The three SC samples with 100% intact SC and estimated 90% and 33% intact SC all showed no difference in *L* or *M* values. The lengths from the *P(r)* analyses were consistent with those from the Guinier analyses (Table 4.1).

4.2.4 Analytical ultracentrifugation analyses of SC and its fragments

The oligomerisation state of SC was determined using sedimentation equilibrium (SE) experiments at concentrations between 0.2 and 2.5 mg/ml, monitored using both interference and absorbance optics (Section 4.4.3). On the assumption that a single species was present, good curve fits with small random residuals were obtained (Figure 4.7). SDS-PAGE analyses before and after data acquisitions confirmed that any cleavage was not detectable. The resulting SC molecular mass values showed a small increase with increasing dilutions from 2.5 mg/ml (lower panels, Figure 4.7). This increase was attributed to an increased electrostatic interaction between SC molecules at higher protein concentrations. Regressions of the dilution series in Figure 4.7 resulted in molecular masses of 79.5 (± 2.9) kDa (interference) and 81.6 (± 5.5) kDa (absorbance) at zero concentration. These values agree well with the sequence-derived molecular weight of 79.6 kDa calculated on the assumption of seven biantennary complex-type oligosaccharides (19.4% by mass) (Figure 4.1a). The assumption of seven tetraantennary oligosaccharides would give an SC molecular mass of 88,800 Da. The SE data therefore showed that SC is a monomer in solution and that all seven putative glycosylation sites are filled. This is consistent with the molecular mass determined by neutron scattering.

The sedimentation coefficient $s^{\circ}_{20,w}$ from sedimentation velocity (SV) experiments monitors the extent of macromolecular elongation, and is independent from the R_G determinations. SV was used to examine both SC and its cleavage into the D1-D3 and D4-D5 fragments as well as an extra sample of cleaved SC with an estimated equimolar mixture of the three, by incubating an SC sample at room-

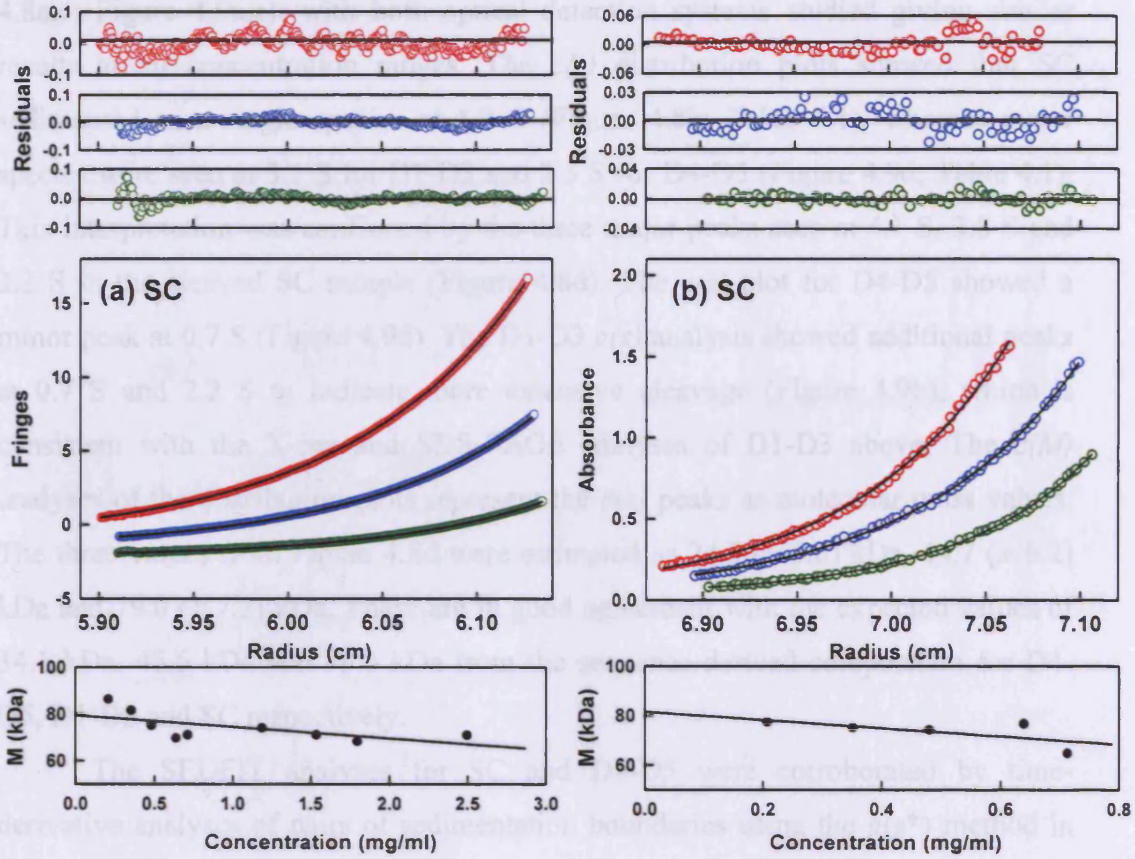

Figure 4.7 Sedimentation equilibrium fits for SC

The curve fits correspond to (a) interference optics and (b) absorbance optics at a wavelength of 280 nm, both at a rotor speed of 14,000 r.p.m. In (a), the concentrations of SC are 2.5 mg/ml (red), 1.2 mg/ml (blue) and 0.48 mg/ml (green). In (b), concentrations are 0.64 mg/ml (red), 0.48 mg/ml (blue), and 0.21 mg/ml (green). Circles represent the experimental data and the continuous lines represent the fits to these. The curve fit residuals are shown in the smaller panels above the exponential fits, using the same colour schemes. The panels below the exponential curve fits show the concentration dependence of the observed molecular mass values.

temperature. The SEDFIT *c(s)* distribution plots identified the $s^{\circ}_{20,w}$ values of the species present in each sample, and are therefore independent of degradation effects. All the calculated and experimental boundary fits are in good agreement (Figure 4.8a,c; Figure 4.9a,c), with both optical detection systems studied giving similar results at all concentration ranges. The *c(s)* distribution plots showed that SC sedimented as a single species at 4.2 S (Figure 4.8b; Table 4.1), whereas major species were seen at 3.1 S for D1-D3 and 2.5 S for D4-D5 (Figure 4.9d; Table 4.1). This interpretation was confirmed by the three major peaks seen at 4.1 S, 3.0 S and 2.2 S in the cleaved SC sample (Figure 4.8d). The *c(s)* plot for D4-D5 showed a minor peak at 0.7 S (Figure 4.9d). The D1-D3 *c(s)* analysis showed additional peaks at 0.7 S and 2.2 S to indicate more extensive cleavage (Figure 4.9b), which is consistent with the X-ray and SDS-PAGE analyses of D1-D3 above. The *c(M)* analyses of the distribution plots represent the *c(s)* peaks as molecular mass values. The three values from Figure 4.8d were estimated as 24.5 (± 5.6) kDa, 44.7 (± 6.2) kDa and 79.0 (± 7.2) kDa. These are in good agreement with the expected values of 34.1 kDa, 45.5 kDa and 79.6 kDa from the sequence-derived composition for D4-D5, D1-D3 and SC respectively.

The SEDFIT analyses for SC and D4-D5 were corroborated by time-derivative analyses of pairs of sedimentation boundaries using the g(s*) method in DCDT+ for SC and D4-D5. Good fits with low residuals were obtained (Figure 4.10). From these, the $s^{\circ}_{20,w}$ values were determined to be 4.0 S and 3.9 S for SC (interference and absorbance optics respectively) and 2.4 S and 2.3 S for D4-D5 (interference and absorbance optics respectively) (Figure 4.10; Table 4.1). The D1-D3 data could not be analysed by DCDT+.

The $s^{\circ}_{20,w}$ values lead to frictional ratios f/f_o, where f_o is the frictional coefficient of the sphere of volume equal to that of the hydrated macromolecule. The f/f_o is a monitor of structural elongation, in which a spherical globular protein will have an f/f_o ratio of 1.0, and many globular proteins show an f/f_o ratio of 1.2 to 1.3. For SC, D1-D3 and D4-D5 the f/f_o ratio were determined to be 1.55, 1.42 and 1.48 respectively (Table 4.1). The similarity of the three f/f_o ratios shows that all three glycoproteins show similar degrees of elongations. As these are consistent with the R_G/R_0 ratios of 1.44-1.49, 1.69 and 1.70 for SC, D1-D3 and D4-D5, they further confirm the folded back structure of SC.

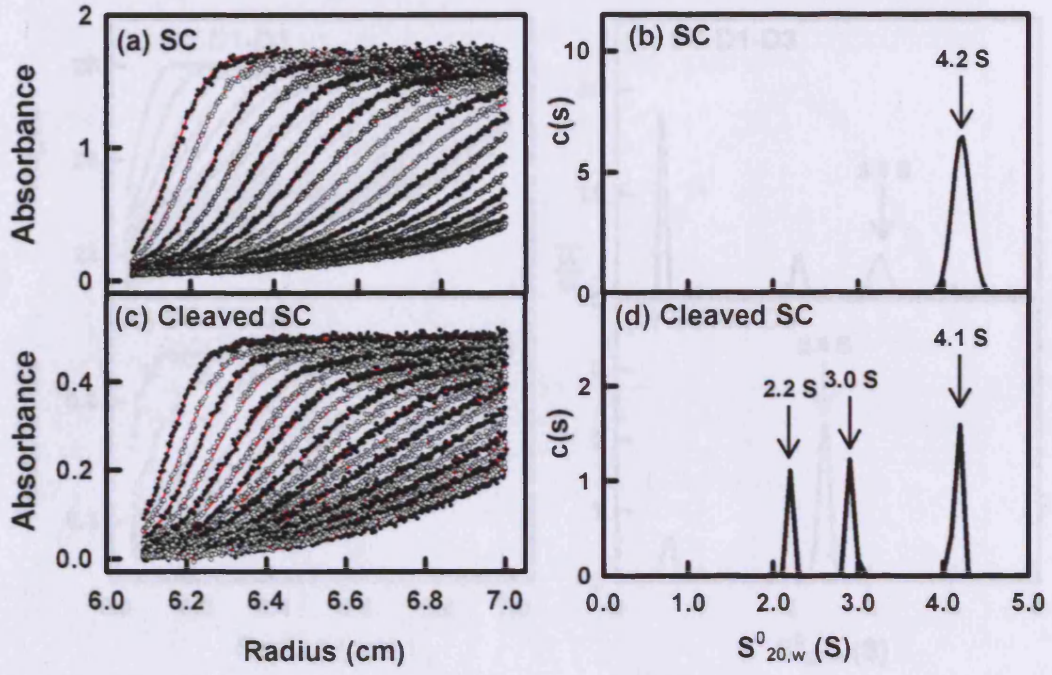

Figure 4.8 SEDFIT sedimentation velocity fits for SC

(a,c) The sedimentation coefficient distribution *c(s)* analyses were performed and the circles represent the experimental data and the continuous red lines represent the fits in the boundary fits. (b,d) The corresponding *c(s)* plots are shown, in which the sedimentation coefficients of the relevant major species are arrowed. (a,b) Absorbance data for SC at a concentration of 1.24 mg/ml and a rotor speed of 30,000 r.p.m., in which only the fits for every fifth scan of the first 100 scans are shown for reason of clarity. (c,d) Absorbance data for a sample of cleaved SC (with estimated 66% intact SC) at a concentration of 0.31 mg/ml and a rotor speed of 42,000 r.p.m., in which every fifth scan of the first 110 scans is shown.

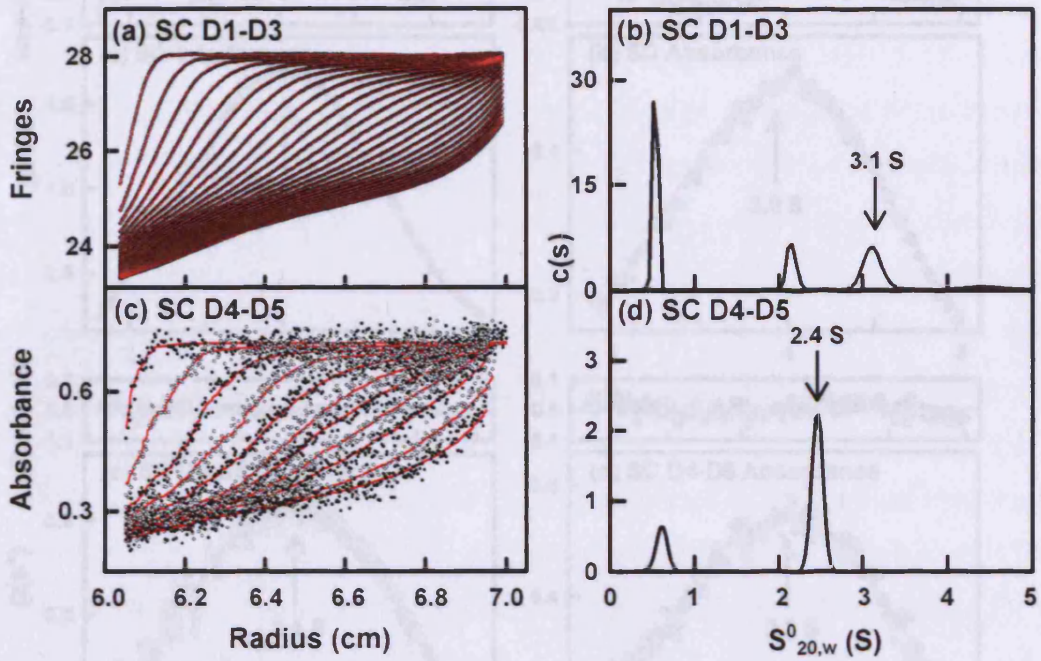

Figure 4.9 SEDFIT sedimentation velocity fits for D1-D3 and D4-D5

(a,c) The sedimentation coefficient distribution *c(s)* analyses were performed and the circles represent the experimental data and the continuous red lines represent the fits. (b,d) The corresponding *c(s)* plots are shown, in which the sedimentation coefficients of the relevant major species are arrowed. (a,b) Interference data for D1-D3 at a concentration of 1.6 mg/ml and a rotor speed of 42,000 r.p.m., in which every sixteenth scan of 500 scans is shown. (c,d) Absorbance data for D4-D5 at a concentration of 0.40 mg/ml and a rotor speed of 42,000 r.p.m., in which every fifth scan of the first 55 scans is shown.

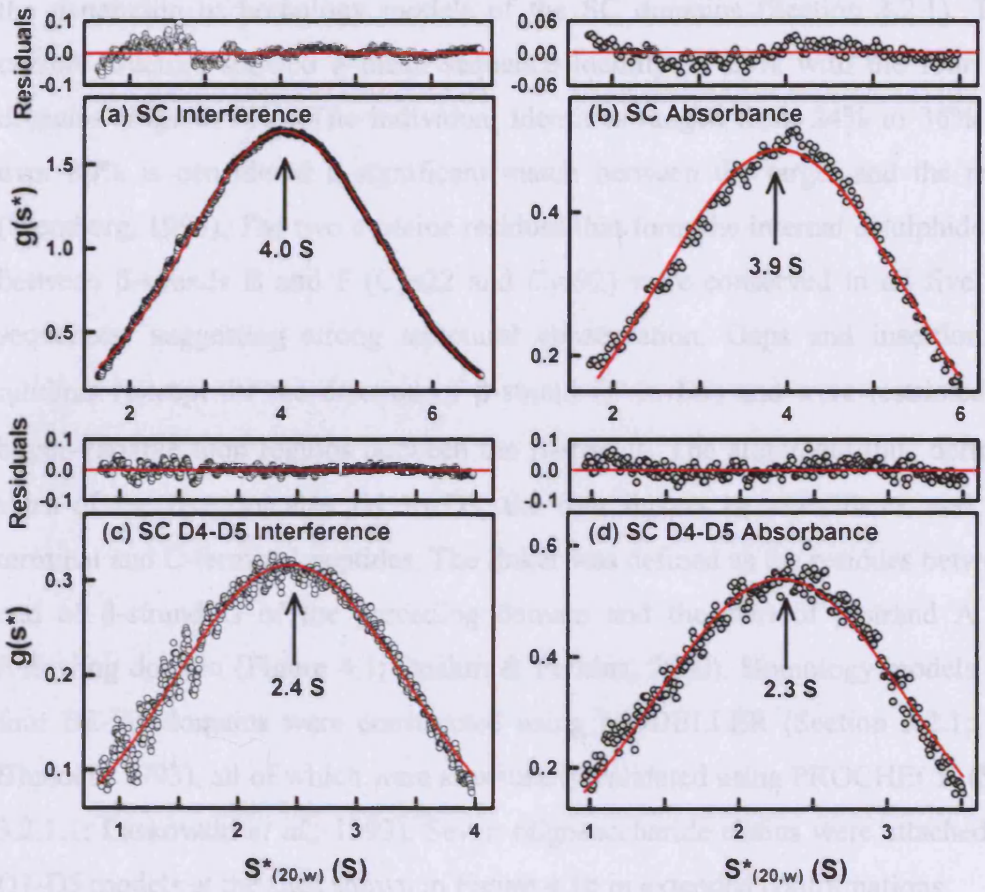

Figure 4.10 DCDT+ sedimentation velocity fits for SC and D4-D5

In the time-derivative method *g(s*)* plots, the experimental data are represented by circles, and the fits are shown by the continuous red lines. The goodness-of-fit residuals are shown above each *g(s*)* analysis. The arrowed $s_{20,w}$ values for SC are (a) 4.0 S from interference data at a concentration of 2.5 mg/ml and (b) 3.9 S from absorbance data at a wavelength of 280 nm at a concentration of 1.2 mg/ml. The corresponding values for D4-D5 are (c) 2.4 S from interference data at 0.73 mg/ml and (d) 2.3 S from absorbance data at 0.40 mg/ml. A rotor speed of 35,000 r.p.m. was used for SC, and 30,000 r.p.m. (interference) and 42,000 r.p.m. (absorbance) for D4-D5.

4.2.5 Constrained modelling of SC and its fragments

The D1 crystal structure of SC to 0.19 nm resolution has a standard two β-sheet sandwich structure composed of the DEBA and A'GFCC'C" β-strands (PDB code 1XED; Hamburger *et al.*, 2004; Figure 1.13). This was used as the template for the generation of homology models of the SC domains (Section 3.2.1). The D1 crystal structure showed a mean sequence identity of 28% with the four D2-D5 domains (Figure 4.1b). The individual identities ranged from 24% to 36%, where over 30% is considered a significant match between the target and the template (Sternberg, 1996). The two cysteine residues that form the internal disulphide bridge between β-strands B and F (Cys22 and Cys92) were conserved in all five D1-D5 sequences, suggesting strong structural conservation. Gaps and insertions were minimal (except for the deletion of β-strand C" in D5) and were restricted to the hyper-variable loop regions between the β-strands. The alignment thus defined the sizes of the five domains D1 to D5, the four linkers between them, and the N-terminal and C-terminal peptides. The linker was defined as the residues between the end of β-strand G of the preceding domain and the start of β-strand A of the following domain (Figure 4.1; Boehm & Perkins, 2000). Homology models for the four D2-D5 domains were constructed using MODELLER (Section 3.2.1; Sali & Blundell, 1993), all of which were structurally validated using PROCHECK (Section 3.2.1.1; Laskowski *et al.*, 1993). Seven oligosaccharide chains were attached to the D1-D5 models at the sites shown in Figure 4.1b in extended conformations.

In order to create full models for intact SC from the D1-D5 domains, randomised conformational libraries of 5,000 unconstrained structures were set up for all four linkers so that these could be used to connect the D1-D5 domains. As the crystal structure of D1 commenced at Ser2, no library was developed for the three residue N-terminal peptide. Lys1 was added to the crystal structure of D1 using MODELLER in order to complete this linker for the SC modelling. Secondary structure analyses of the 42 residue C-terminal peptide predicted that there is little secondary structure (Rost *et al.*, 2003). BLASTP sequence searches of over 39,000 structures in the PDB (September 2006) did not reveal any significant similarity with a structural template for the C-terminal peptide (Schaffer *et al.*, 2001). The amino acid composition showed that there is a high proportion of 12 negatively-charged Asp/Glu residues and nine positively-charged Arg/Lys residues in this sequence,

together with 14 small Gly/Ala/Ser residues, all of which promotes random structures (Sun *et al.*, 2004). Accordingly the C-terminal peptide was modelled in the same way as the four linkers as another unconstrained structure.

The modelling of the smallest SC fragment, D4-D5 was used to initiate the SC modelling for reason of simplicity. The D4-D5 linker peptide and the C-terminal peptide were the only conformational variables. Five trial-and-error searches were carried out in which the assumptions made for these two conformations were varied. The initial use of unconstrained D4-D5 and C-terminal linkers in Search 1 resulted in 5000 D4-D5 models, where many were too compact and showed steric overlaps between D4 and D5 (blue filled spheres in Figure 4.11d,e,f). The Search 1 D4-D5 models gave R_G values ranging from 2.10 nm to 3.07 nm, which were too low compared to the experimental D4-D5 R_G value of 3.16 nm. Accordingly, Searches 2 to 5 each generated a linker library of 5000 randomised conformations for the D4-D5 linker that were constrained to be between 95%-100% of their maximum allowed length, but which were left free to reorientate. The maximum peptide bond length was 0.35 nm. In each of Search 2, 3, 4 and 5, the C-terminal peptide was set to be a minimum length of 5 nm, 7 nm, 9.8 nm and 11.8 nm respectively in order to enforce the generation of a wide range of conformations for this. The combination of Searches 2 to 5 gave 6000 models that spanned a wide range of conformational separations of the N-terminus and C-terminus from 0.6 nm to 20.6 nm as required (Figure 4.11f). The best-fit models were identified from the 6,000 models using filters. Filtering to retain a minimum of 346 hydrated spheres (i.e. retaining only those models that have no significantly overlapping regions) and a R_G value between 3.00-3.20 nm gave 1011 satisfactory models (Table 4.2). Sorting the 1011 models using the R factor goodness-of-fit parameter identified the best nine models (yellow and red in Figure 4.11d,e,f). The best D4-D5 model with an R_G value of 3.15 nm and a R_{XS} value of 1.14 nm (red in Figure 4.11d,e,f) was then selected as this gave the closest agreement with the experimental values of 3.16 ± 0.04 nm and 1.11 ± 0.08 nm respectively. The visual fit of the best-fit D4-D5 calculated $I(Q)$ curve with the experimental $I(Q)$ scattering curve in Figure 4.12d is excellent. The sedimentation coefficient of 2.04 S from this best-fit model is close to the experimental values of 2.3 S to 2.4 S (Table 4.2). All nine best-fit D4-D5 models showed that the D4 and D5 domains were V-shaped with an extended C-terminal peptide.

Glycoprotein	Filter	Models	Spheres	R_G (nm)	R_{XS} (nm)	R factor (%)	N-C terminal distance (nm)	$s^*_{20,w}$ (S)
SC (X-ray)	None	5000	678-869	3.43-4.71	1.05-2.00	3.3-19.4	2.2-22.3	n.a.
	R_G	268	680-865	3.46-3.75	1.40-2.00	3.4-7.4	3.1-18.3	n.a.
	R_G, R_{XS}, N	10	824-865	3.53-3.74	1.71-1.88	3.9-7.4	4.7-14.4	3.95-4.24
	Best-fit	1	850	3.53	1.80	4.4	7.2	4.06
	CHO in	1	853	3.52	2.00	5.8	7.2	4.09
	Linear	1	1125	4.72	1.42	18.0	21.5	3.38
	Experimental	n.a.	830	3.53 ± 0.03	1.76 ± 0.08	n.a.	n.a.	3.9-4.0
SC (neutron)	None	5000	498-625	3.17-4.37	1.01-1.77	3.8-11.2	2.2-22.3	n.a.
	R_G	268		3.22-3.52	1.24-1.77	3.9-7.1	3.1-18.3	n.a.
	R_G, R_{XS}, N	10	592-630	3.30-3.51	1.58-1.74	5.9-7.1	4.7-14.4	n.a.
	Best-fit	1	619	3.30	1.70	6.8	7.2	n.a.
	CHO in	1	616	3.30	1.77	7.4	7.2	n.a.
	Linear model	1	850	4.38	1.33	10.7	21.5	n.a.
	Experimental	n.a.	626	3.63 ± 0.28	1.30 ± 0.10	n.a.	n.a.	n.a.
D1-D3 (X-ray)	None	1000	421-525	2.47-3.68	1.19-1.77	13.2-16.8	4.6-15.1	n.a.
	R_G	65	474-518	3.40-3.49	1.24-1.44	13.2-16.5	10.0-14.8	n.a.
	R_G, R_{XS}	10	475-509	3.40-3.49	1.24-1.30	13.2-15.7	10.2-13.3	2.64-3.02
	Best linear	1	484	3.40	1.27	13.2	11.7	3.02
	CHO in	1	484	3.26	1.24	11.6	11.7	3.00
	Bent model	1	507	3.11	1.66	15.8	6.6	3.04
	Best hybrid	1	n.a.	2.98	1.06	6.0	n.a.	n.a.
	D1 domain	1	179	1.90	0.73	19.9	2.9	1.64
	Experimental	n.a.	474	3.44 ± 0.27	1.07 ± 0.07	n.a.	n.a.	3.1
D4-D5 (X-ray)	None	6000	278-410	2.02-3.71	0.90-1.76	7.3-18.0	0.6-20.6	n.a.
	R_G, N	1011	346-410	3.00-3.20	0.98-1.68	7.7-17.2	5.3-19.1	n.a.
	R_G, N, R-factor	9	346-368	3.06-3.19	1.10-1.22	7.7-8.1	10.5-17.7	2.00-2.26
	Best-fit	1	367	3.15	1.14	8.1	11.6	2.04
	CHO in	1	367	3.13	0.86	8.6	11.6	2.22
	Linear model	1	384	3.71	1.45	16.4	18.5	1.94
	Experimental	n.a.	356	3.16 ± 0.04	1.11 ± 0.08	n.a.	n.a.	2.3-2.4

Table 4.2 Summary of the X-ray and neutron modelling fits for the SC, D1-D3 and D4-D5 solution structures

n.a. - not available

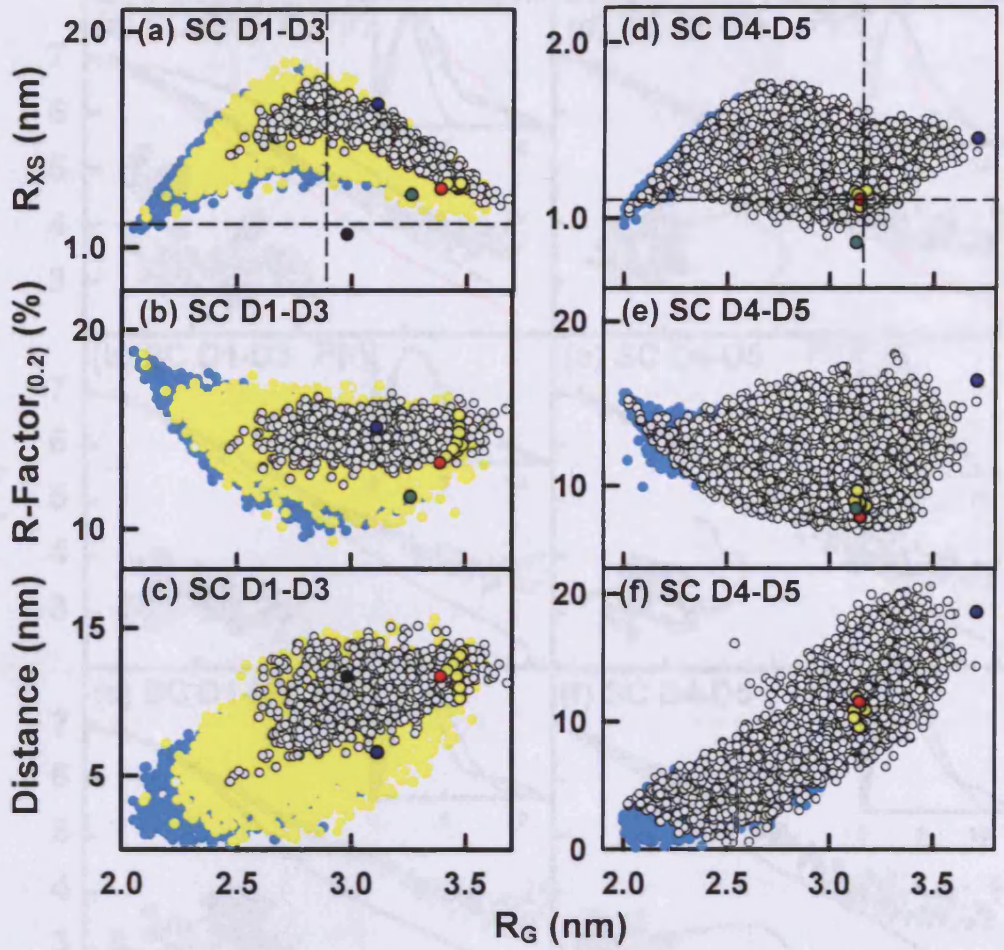

Figure 4.11 Constrained scattering analyses of the X-ray and neutron data for D1-D3 and D4-D5

(a,b,c) The 5000 D1-D3 models represented by blue circles show the outcome of all linkers being unrestrained and the 3000 models represented by yellow circled show the linker restrained between 95-100% of their maximum length. The final analysis, where 1000 models where only the D1-D2 linker was varied is shown by open circles. (d,e,f) The 5000 D4-D5 models where all linkers were unconstrained and varied are shown by blue circles. The 6000 models where the D4-D5 linker and the C-terminal peptide are constrained are shown by open circles. The ten best-fit models (●) and within them the best-fit model (●) and the representative poor-fit model (●) are highlighted. The best-fit models with compact oligosaccharide conformations (●) and the best-fit hybrid D1-D3 model (●) are shown. Panels (a, d) compare the R_{XS} values with the R_G values. Panels (b, e) compare the R factor values with the R_G values. Panels (c, f) compare the distance between the N-terminal and C-terminal Cα residues with the R_G values. The dashed lines correspond to the experimental R_G and R_{XS} values in each case as appropriate.

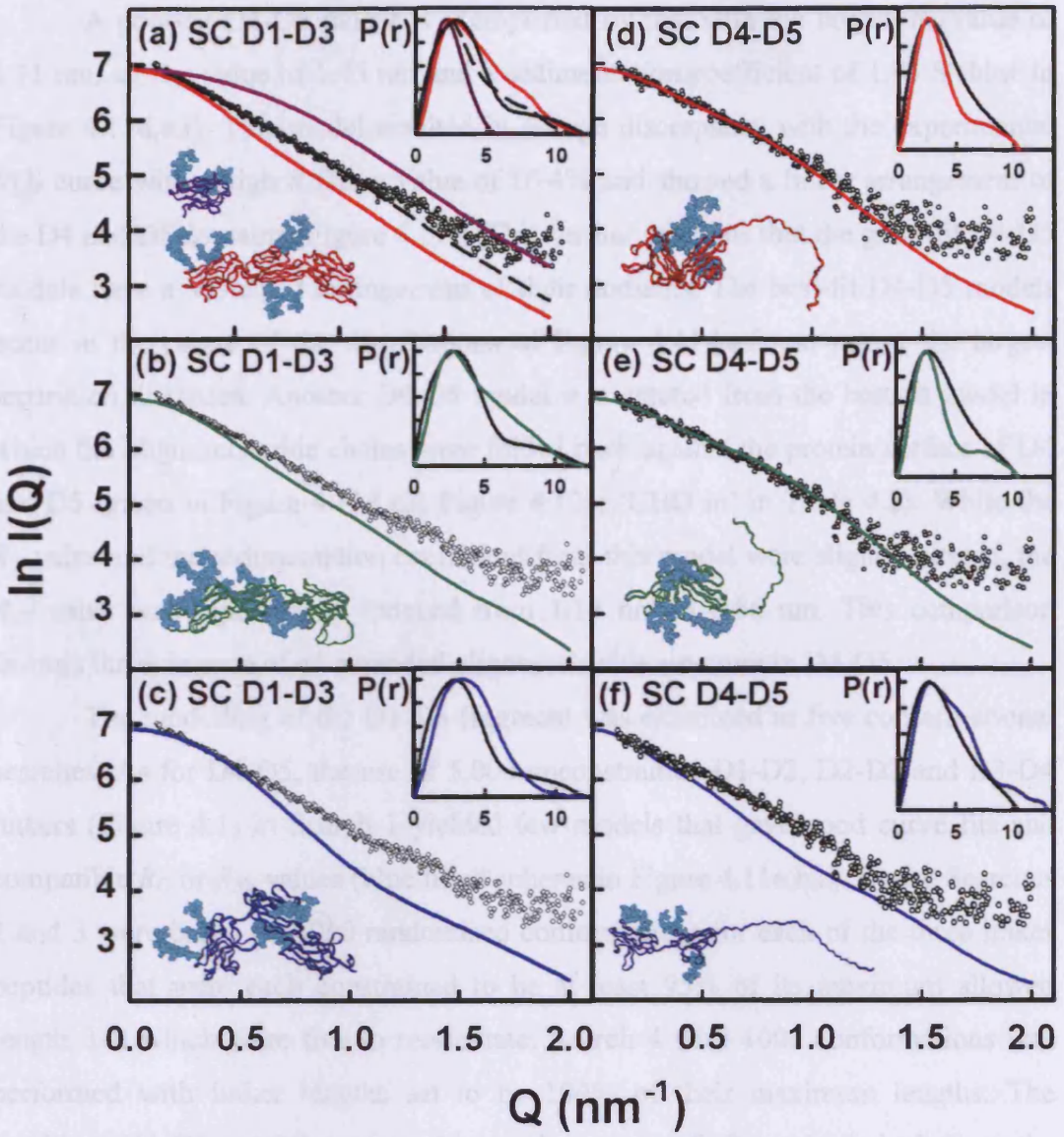

Figure 4.12 Best-fit and poor-fit D1-D3 and D4-D5 models for the X-ray and neutron data

The colour coding of the models corresponds to that of Figure 4.11. The open circles denote the experimental scattering curves, while the continuous lines show the modelled curves. A ribbon trace of the model used in each fit is shown. The modelled *P(r)* curves are shown as insets to the right, with the experimental curve in black and the modelled curve in colour. (a) The best-fit hybrid D1-D3 curve is a dashed line. A ribbon trace of the model used in each fit is shown, with the oligosaccharides in cyan. (a,d) The best fit models in red, (b,e) the best-fit models with compacted oligosaccharide conformations in green and (c,f) the representative poor-fit models in blue are shown.

A poor-fit D4-D5 model is exemplified by that with the largest R_G value of 3.71 nm, an R_{XS} value of 1.45 nm and a sedimentation coefficient of 1.94 S (blue in Figure 4.11d,e,f). This model resulted in a large discrepancy with the experimental *I(Q)* curve with a high *R* factor value of 16.4% and showed a linear arrangement of the D4 and D5 domains (Figure 4.12f). This further confirms that the good-fit D4-D5 models have a V-shaped arrangement of their domains. The best-fit D4-D5 models occur in the centre of the distributions of Figure 4.11d,e,f and not at the largest separation distances. Another D4-D5 model was created from the best-fit model in which the oligosaccharide chains were folded back against the protein surface of D4 and D5 (green in Figure 4.11d,e,f; Figure 4.12e; 'CHO in' in Table 4.2). While the R_G value and the sedimentation coefficient from this model were slightly altered, the R_{XS} value was significantly reduced from 1.14 nm to 0.86 nm. This comparison favours the existence of an extended oligosaccharide structure in D4-D5.

The modelling of the D1-D3 fragment was examined in five conformational searches. As for D4-D5, the use of 5,000 unconstrained D1-D2, D2-D3 and D3-D4 linkers (Figure 4.1) in Search 1 yielded few models that gave good curve fits and compatible R_G or R_{XS} values (blue filled spheres in Figure 4.11a,b,c). Hence Searches 2 and 3 were based on 1000 randomised conformations for each of the three linker peptides that were each constrained to be at least 95% of its maximum allowed length, but which were free to reorientate. Search 4 with 1000 conformations was performed with linker lengths set to be 100% of their maximum lengths. The resulting D1-D3 models spanned a wide range of R_G values including the experimental value, but all of these gave R_{XS} values that were too high compared to the experimental R_{XS} value (yellow filled spheres in Figure 4.11a,b,c). The D2-D3 linker is only one amino acid in length (Figure 4.1). For this reason, any movement in this linker resulted in steric overlap between the D2 and D3 domains. Hence in Search 5, the D2 and D3 domains were joined by a short linker in a single extended and fixed conformation. The C-terminal D3-D4 peptide was also fixed in one extended conformation. Only the D1-D2 linker was varied in Search 5, and as shown in Table 4.2, whilst 65 models show compatible R_G values to the experimental value, none show the expected R_{XS} value of 1.07 nm. The best-fit D1-D3 model with the lowest *R* factor was determined, yet the *R* factor value was high at 13.2% (red in Figure 4.11a,b,c; Table 4.2). All ten best-fit D1-D3 models exhibited linear domain arrangements (yellow in Figure 4.11a,b,c; Figure 4.12a). An alternative D1-D3

model in which the extended oligosaccharide conformations were folded back against the protein surface was also unable to account for the reduced R_{XS} value (green in Figure 4.11a,b,c; Figure 4.12b).

Good D1-D3 curve fits were finally obtained by postulating that the observed X-ray scattering curve is a mixture of three-domain and single-domain fragments. This heterogeneity had been identified by SDS-PAGE and the *c(s)* sedimentation analyses (Figure 4.09b). Composite scattering curves were computed from the extended D1-D3 model that gave the lowest R factor of 13.2% in Search 5 and the single D1 model on its own. The sum of 74 % of the D1-D3 curve and 26 % of the D1 curve gave the best curve fit with a good R factor of 6.0 % (black in Figure 4.11a,b,c; Figure 4.12a). Other poor-fit D1-D3 models were tested to confirm this interpretation. For example, the scattering curve of a U-shaped D1-D3 model (blue in Figure 4.11a,b,c; Figure 4.12c) similarly combined with the D1 model scattering curve were unsuccessful in reducing the R factor or the R_{XS} value.

The modelling of intact SC was performed using two different approaches. Search 1 was based on the use of 5000 models made from all five unconstrained linkers (blue filled circles in Figure 4.13). Convex distributions of the R_{XS} values as a function of the R_G values were observed for both the X-ray and neutron modelling fits, where the maximum agreed well with the observed R_G values of 3.53 nm (X-rays) and 3.63 nm (neutrons). As desired, the R factor values showed minima at the observed R_G values in Figure 4.13b,e. However due to the unconstrained nature of this search many of these models showed steric overlap, thus a second approach was employed. Search 2 utilised the two best fit models of D1-D3 and D4-D5 and only varied the long D3-D4 linker (Figure 4.1). The distributions of the unfilled circles in Figure 4.13 showed that, while these followed the same trends as the cyan circles of the unconstrained Search 1, Search 2 resulted in fewer sterically overlapping models with low R_G values. The best SC structures were identified by filtering based on the extent of agreement with the R_G and R_{XS} values and the lowest R factor values (Table 4.2). Thus 268 SC models agree well with the experimental X-ray R_G value of 3.53 nm, of which 70 models agree well with the experimental R_{XS} value of 1.76 nm. The removal of the SC models that showed partially overlapping domain structures and selection of the models with best R factor values then gave the 10 best-fit SC models (yellow in Figure 4.13; Table 4.2). Search 2 showed that all 10 best SC models showed a folded-back domain arrangement in which the D1-D3 domains are in

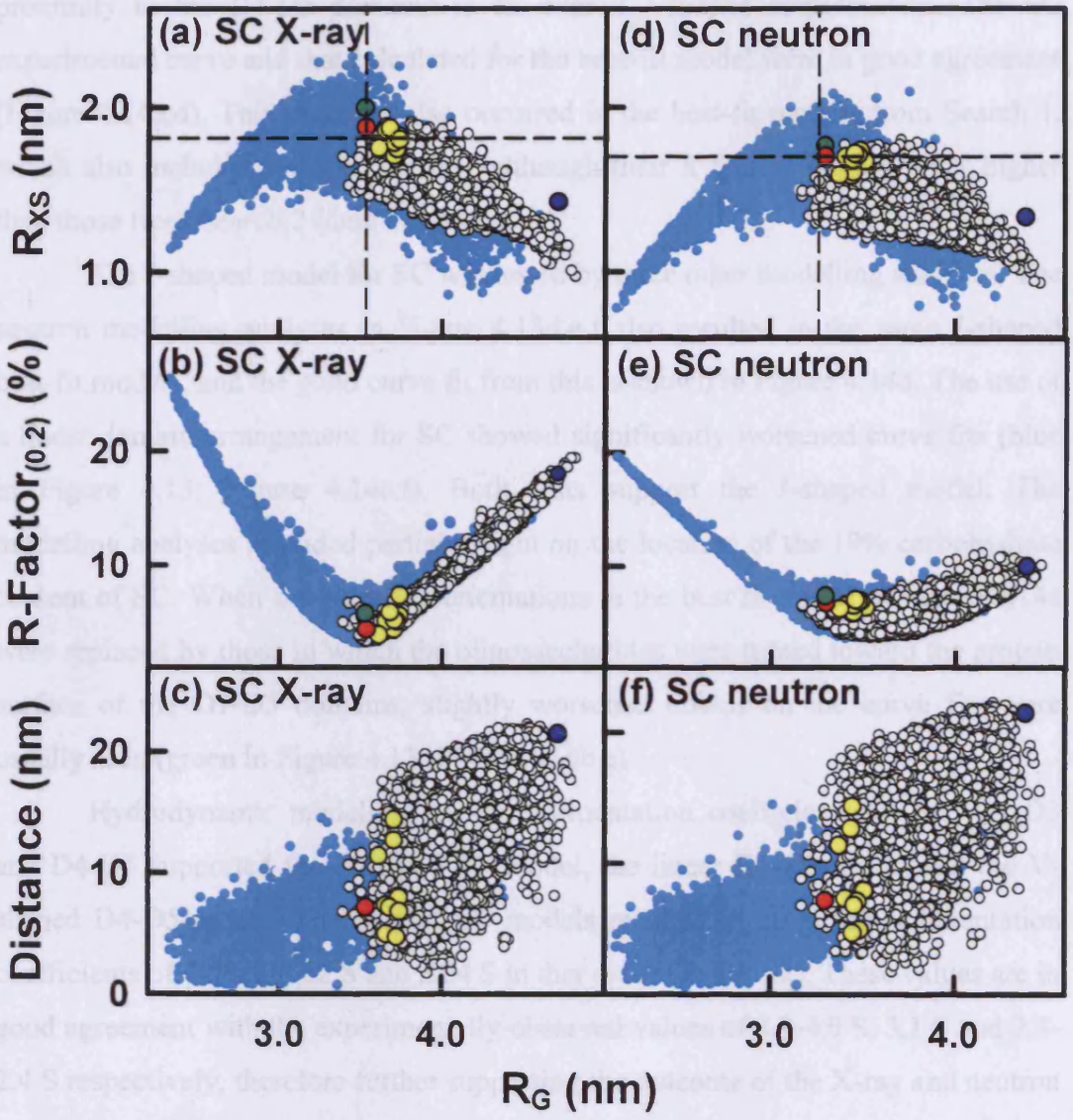

Figure 4.13 Constrained scattering analyses of the SC X-ray and neutron data
The 5,000 SC models represented by blue circles show the outcome of all linkers being unrestrained in the generation of the X-ray and neutron SC models. The final analysis, where just the D3-D4 linker was varied and D1-D3 and D4-D5 was held fixed is represented by 5,000 unfilled circles. The ten best-fit models (●) and the best-fit SC model (●) are shown. Separately from this, the best-fit model with compact oligosaccharide conformation (●) and the representative poor-fit SC model (●) are shown. Panels (a, d) compare the R_{XS} values with the R_G values. Panels (b, e) compare the R factor values with the R_G values. Panels (c, f) compare the distance between the N-terminal and C-terminal Cα residues in SC with the R_G values. The dashed lines correspond to the experimental R_G and R_{XS} values in each case as appropriate.

proximity to the D4-D5 domains in an overall J-shaped conformation. The SC experimental curve and that calculated for the best-fit model were in good agreement (Figure 4.14a,d). This outcome also occurred in the best-fit models from Search 1, which also included J-shaped models, although their *R* factors were slightly higher than those from Search 2 (data not shown).

The J-shaped model for SC was tested by three other modelling analyses. The neutron modelling analyses in Figure 4.13d,e,f also resulted in the same J-shaped best-fit models, and the good curve fit from this is shown in Figure 4.14d. The use of a linear domain arrangement for SC showed significantly worsened curve fits (blue in Figure 4.13; Figure 4.14c,f). Both tests support the J-shaped model. The modelling analyses provided partial insight on the location of the 19% carbohydrate content of SC. When the extended orientations in the best fit model of Figure 4.14a were replaced by those in which the oligosaccharides were turned toward the protein surface of the D1-D5 domains, slightly worsened effects on the curve fits were usually seen (green in Figure 4.13; Figure 4.14b,e).

Hydrodynamic modelling of the sedimentation coefficients for SC, D1-D3 and D4-D5 supported the J-shaped SC model, the linear D1-D3 model and the V-shaped D4-D5 model. The best-fit SC models resulted in predicted sedimentation coefficients of 4.06 S, 3.02 S and 2.04 S in that order (Table 4.2). These values are in good agreement with the experimentally-observed values of 3.9-4.0 S, 3.1 S and 2.3-2.4 S respectively, therefore further supporting the outcome of the X-ray and neutron modelling of SC.

These results of SC modelling are summarised in the form of a comparison of SC with IgA1 and IgA2 monomers (Figure 4.15), the superimposition of the ten best-fit SC models (Figure 4.16), a stereo view of the best-fit model with the carbohydrate chains shown (Figure 4.17a), an electrostatic map of D1-D3 and D4-D5 (Figure 4.17b) and a view showing the availability of dIgA binding motifs within SC (Figure 4.17c).

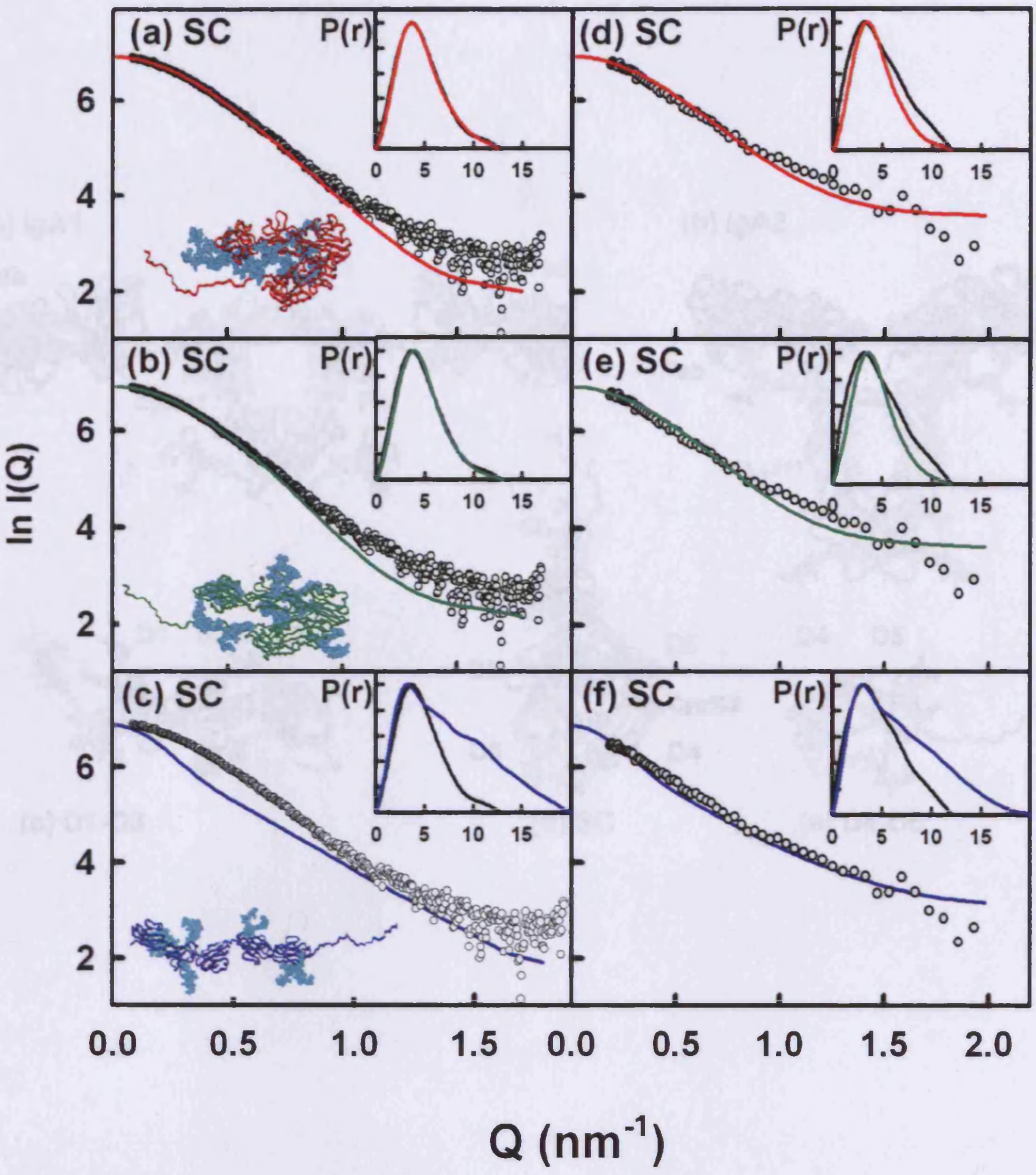

Figure 4.14 Best-fit and poor-fit SC models for the X-ray and neutron data

The colour coding of the models corresponds to that in Figure 4.13. The open circles denote the experimental scattering curves, while the continuous lines show the modelled curves. A ribbon trace of the model used in each fit is shown. The modelled *P(r)* curves are shown as insets to the right, with the experimental curve in black and the modelled curve in colour. Panels (a, d) show in red the best-fit model. Panels (b, e) show in green the best-fit model with folded-back oligosaccharide chains. Panels (c, f) show in dark blue the extended poor-fit model that gives poor curve fits to the experimental scattering curves. The carbohydrate chains in each model are shown in cyan.

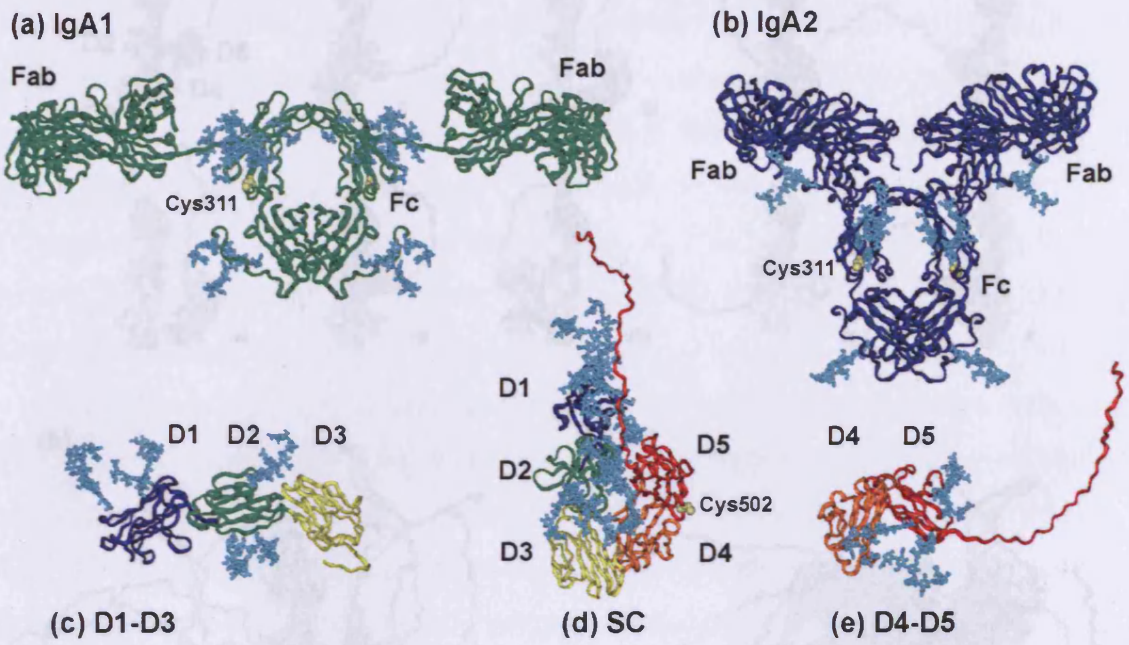

Figure 4.15 Comparison of the solution scattering models for monomeric IgA1 and IgA2 with the best-fit models of SC and its D1-D3 and D4-D5 fragments

(a,b) The models for IgA1 and IgA2 are shown in green and blue respectively, in which the Cys311 are highlighted in yellow. (c,d,e) The best-fit D1-D3, SC and D4-D5 models are shown in the same colouring scheme used in Figure 4.1a for the D1-D5 domains. In the SC model, the covalent contact residue Cys502 on D5 is coloured yellow. The carbohydrate chains in each model are shown in cyan.

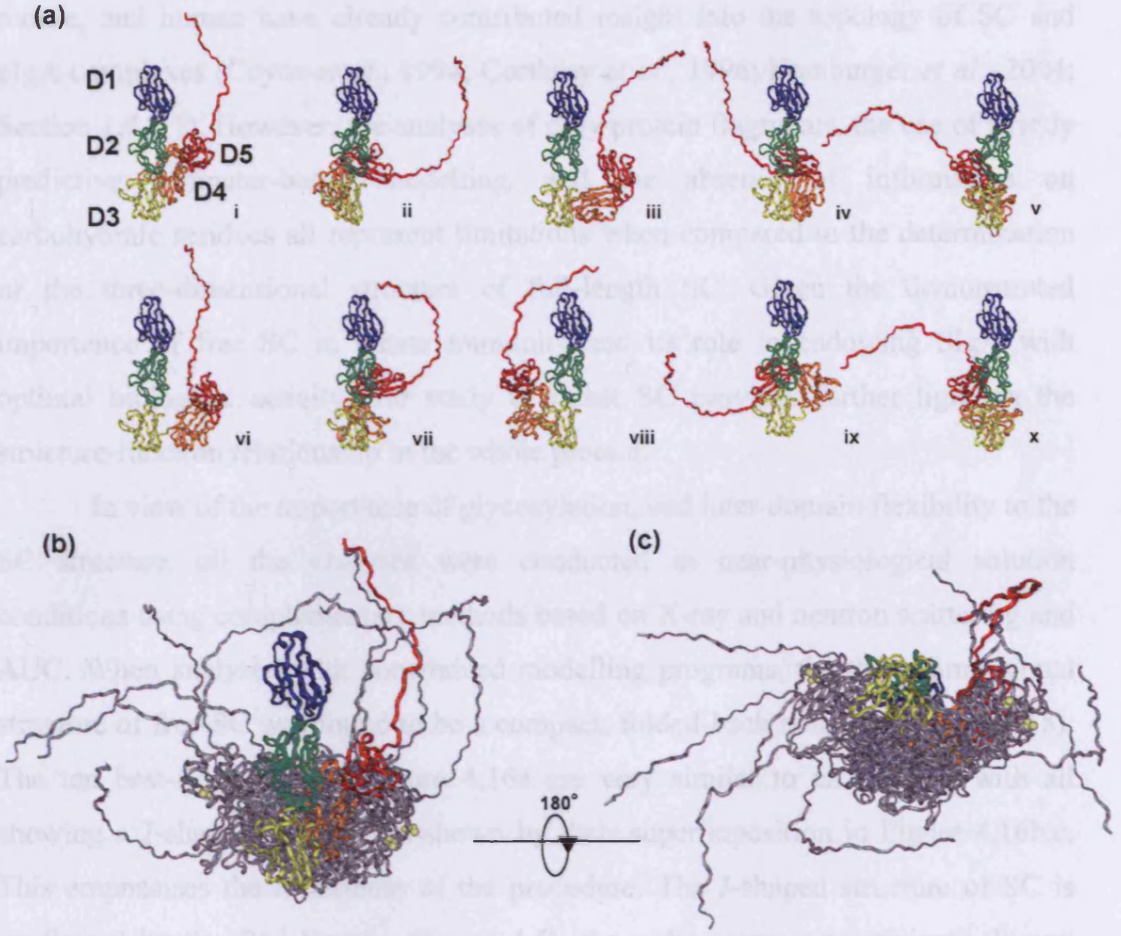

Figure 4.16 Ten best-fit models of SC

(a) The ten best-fit models (● in Figure 4.13) are shown with the D1-D3 domains shown in the same orientation in all ten models (i-x). Model (i) is the best-fit model from Figure 4.14a. The five domains are coloured from blue (N-terminus) to red (C-terminus) to follow that of Figure 4.1a. (b,c) Two orthogonal views of the ten best-fit SC structures superimposed. The best-fit model from part a(i) is shown in the same colour scheme and the other nine models are shown in grey. Residue Cys502 in D5 that covalently binds to IgA Cys311 in SIgA is coloured yellow.

4.3 Discussion

Molecular modelling and crystal structures of SC domains from rabbit, mouse, and human have already contributed insight into the topology of SC and pIgA complexes (Coyne *et al.*, 1994; Corthésy *et al.*, 1996; Hamburger *et al.*, 2004; Section 1.4.2.1). However, the analyses of only protein fragments, the use of strictly predictive computer-based modelling, and the absence of information on carbohydrate residues all represent limitations when compared to the determination of the three-dimensional structure of full-length SC. Given the demonstrated importance of free SC in innate immunity and its role in endowing SIgA with optimal biological activity, our study of intact SC provides further light on the structure-function relationship in the whole protein.

In view of the importance of glycosylation, and inter-domain flexibility to the SC structure, all the analyses were conducted in near-physiological solution conditions using complementary methods based on X-ray and neutron scattering and AUC. When analysed with constrained modelling programs, the three-dimensional structure of free SC was found to be a compact, folded-back structure (Figure 4.15). The ten best-fit models of Figure 4.16a are very similar to each other, with all showing a J-shaped structure as shown by their superimposition in Figure 4.16b,c. This emphasises the robustness of the procedure. The J-shaped structure of SC is confirmed by the *P(r)* lengths (Figure 4.5), the sedimentation coefficients (Figure 4.8) and the constrained modelling (Figure 4.13; Figure 4.14). The point at which SC is folded back to form a compact J-shaped structure was located to the junction between the D3 and D4 domains by two different modelling searches of all five SC domains. The linker between D3 and D4 is unusually long in SC, being ten residues in length (Figure 4.1). In support of this model, the length and sedimentation coefficient data obtained with the D1-D3 and D4-D5 fragments both fit perfectly with the SC structure (Figure 4.6; Figure 4.9; Figure 4.11; Figure 4.12), and reinforce the accuracy of the SC model.

This compact, folded-back structure of SC was unexpected in the light of the extended domain arrangements of the structurally-related seven Ig-like domains in carcinoembryonic antigen and four FnIII domains in anosmin-1 (Boehm & Perkins, 2000; Hu *et al.*, 2005). In fact, the linker between D3 and D4 comprises 10 amino acids, and is unusually long (Figure 4.1). This length provides the requisite flexibility needed to form the "J-shaped" conformation. It is noteworthy that the

cleavage site between Arg336 and Ser337 is exposed on the surface of SC in all ten best-fit SC structures (Figure 4.16), thus accounting for its susceptibility to proteolytic cleavage, and justifying the appearance of the two D1-D3 and D4-D5 polypeptides. The Poisson-Boltzmann electrostatic maps of the carbohydrate-free surfaces in D1-D3 and D4-D5 (Figure 4.17b) suggest that there are spatially well-defined areas of negative and positive charges respectively on these which can attract each other to form the bent structure of SC. The short linker between D2-D3 may act to reduce flexibility between D2 and D3, hence facilitating the binding of D4 and D5 to these.

The modelling showed that the 42 residue C-terminal peptide connecting SC with the transmembrane domain of the pIgR exhibits no particular folding (Figure 4.1; Figure 4.16). This suggests that D1-D5 of SC can freely move with respect to this C-terminal linker, thus enhancing their likelihood to encounter and interact with their cargo represented by polymeric immunoglobulins present in the lamina propria. This length and lack of secondary structure of the C-terminal peptide facilitates access to proteases in order that unbound SC or SIgA can be released at mucosal surfaces (Figure 1.11). The unfolded nature of the linker also suggests that the structure of SC identified in this work will be similar to that when incorporated in its pIgR precursor, as steric interference with components on the cell membrane will be limited.

The seven Asn residues that constitute the attachment points for carbohydrate residues are located on the same side of SC, implying that these carbohydrate moieties are clustered within a restricted part of the SC structure (Figure 4.17a). The curve fits showed a preference for modelling the seven oligosaccharide chains in conformations extended from the protein surface (Figure 4.13; Figure 4.14a,b). Their precise spatial locations remain unassigned by constrained modelling, and it is likely that most of the oligosaccharides are free to interact with multiple target structures. A preferential distribution of oligosaccharides on one face of SC can contribute to tether free and/or IgA-bound SC efficiently to antigenic structures found at the surface of microorganisms (Perrier *et al.*, 2006).

In the context of the interaction between SC and dIgA, the location of the oligosaccharides on one face of SC leave uncovered, and thus accessible, the three CDR-like motifs present in D1, as well as Cys502 in D5 (Figure 4.17c). These constitute two critical elements for the interaction with dIgA (Bakos *et al.*, 1994;

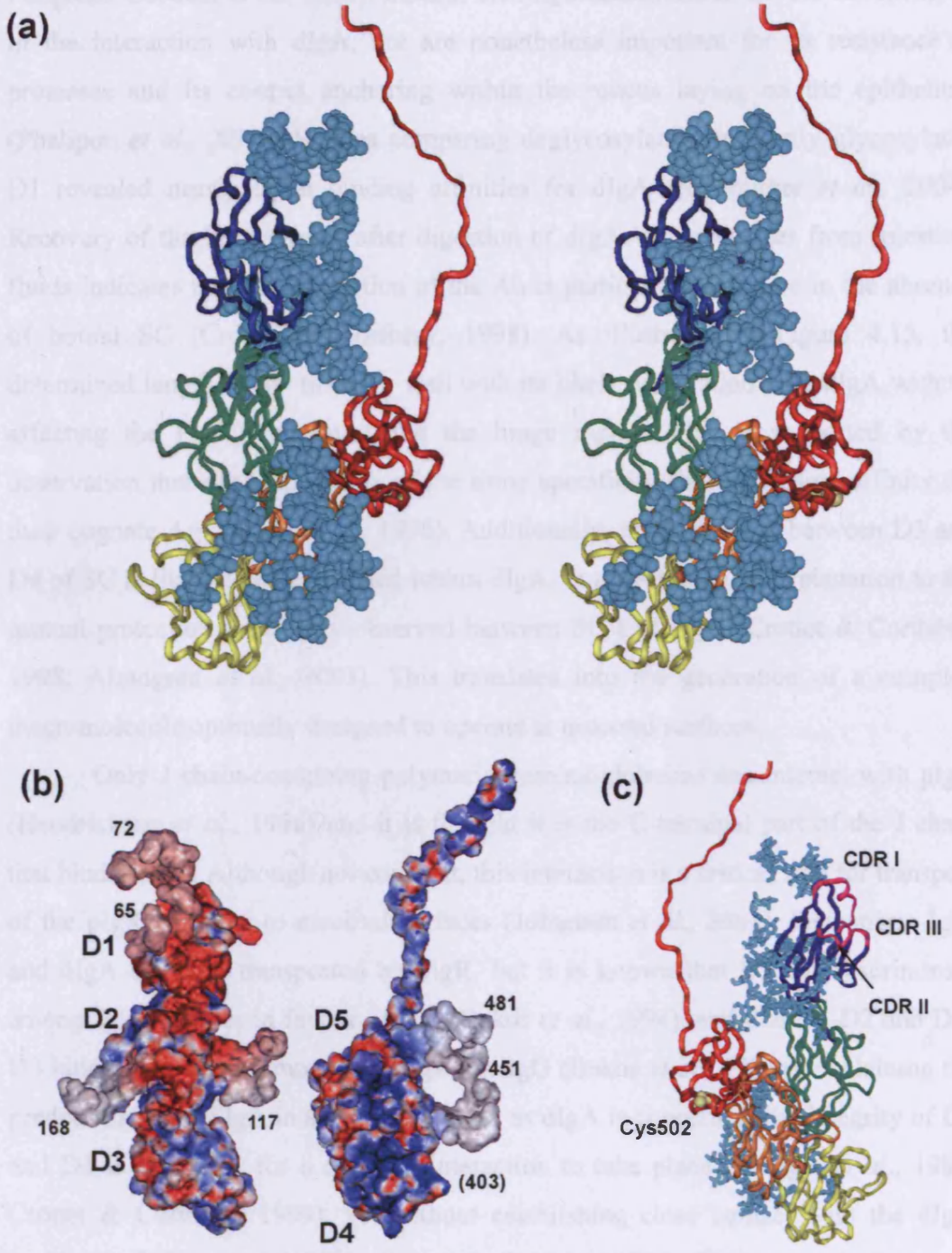

Figure 4.17 Detailed views of the best-fit model of SC

D1-D5 are coloured from blue (N-terminus) to red (C-terminus) to follow that of Figure 4.1a. (a) Stereo view of the best fit model of Figure 4.13a. (b) Electrostatic views of the continuous protein surfaces seen in the D1-D3 and D4-D5 models Red: acidic; blue: basic. The oligosaccharide chains are numbered to follow that of Figure 4.1a. (c) Structure of the best-fit SC model rotated by 180° highlighting SC Cys502 (yellow) and the three CDRs (pink) in D1.

Fallgreen-Gebauer *et al.*, 1993). Indeed, the oligosaccharides of SC are not involved in the interaction with dIgA, but are nonetheless important for its resistance to proteases and its correct anchoring within the mucus laying on the epithelium (Phalipon *et al.*, 2002). Studies comparing deglycosylated and partly glycosylated D1 revealed nearly equal binding affinities for dIgA (Hamburger *et al.*, 2004). Recovery of the Fd fragment after digestion of dIgA with proteases from intestinal fluids indicates that the Fc portion of the Ab is particularly sensitive in the absence of bound SC (Crottet & Corthésy, 1998). As illustrated in Figure 4.15, the determined length of SC fits very well with its likely association with dIgA without affecting the $F(ab')_2$ structure and the hinge region. This is supported by the observation that dIgA and SIgA of the same specificity show the same affinity for their cognate Ag (Lüllau *et al.*, 1996). Additionally, the labile link between D3 and D4 of SC is likely to be concealed within dIgA, thus providing an explanation to the mutual protection previously observed between SC and dIgA (Crottet & Corthésy, 1998; Almogren *et al.*, 2003). This translates into the generation of a complex macromolecule optimally designed to operate at mucosal surfaces.

Only J chain-containing polymeric immunoglobulins can interact with pIgR (Hendrickson *et al.*, 1996) and it is thought it is the C-terminal part of the J chain that binds to SC. Although not covalent, this interaction is a critical step for transport of the pIgR complex to mucosal surfaces (Johansen *et al.*, 2001). Pentameric IgM and dIgA are both transported by pIgR, but it is known that D1 can discriminate among Ig subclasses in favour of IgA (Bakos *et al.*, 1994), and that D1-D2 and D1-D3 bind dIgA, but not monomeric IgA nor IgG (Bakos *et al.*, 1991a), explaining the predominance of SIgA in secretions. As far as dIgA is concerned, the integrity of D2 and D3 is necessary for a covalent interaction to take place (Frütiger *et al.*, 1987; Crottet & Corthésy, 1999), yet without establishing close contact with the dIgA backbone (Bakos *et al.*, 1991b; Crottet & Corthésy, 1999). This argues in favour of the "zipper effect" model initiated by the interaction of D1 with one Cα chain of pIgA, progression to the J chain and establishment of the disulphide bond between Cys311 of the Cα2 domain of IgA and Cys502 of D5 of SC (Fallgreen-Gebauer *et al.*, 1993).). In the J-shaped structure of SC (Figure 4.16; Figure 4.17), D1 remains appropriately exposed to initiate this "zipper" binding interaction.

Several unsolved issues remain, such as the identification of the residues in SC that are involved in the interactions with the J chain and with the residues 402-410 of the Cα3 domain in the Fc fragment of IgA (Hexham *et al.*, 1999). Likewise, even though the CDR-like motifs of D1 have been extensively described (Hamburger *et al.*, 2004), their ligands on the dIgA molecule remain unidentified. In this respect, based on electron micrographs of human dIgA displaying a partly bent tail-to-tail arrangement (Figure 1.9; Dourmashkin *et al.*, 1971), one can envisage that the Fc domain spacing in dIgA is suitable for the binding of one SC molecule (Figure 4.15). Whether the compact conformation of SC opens up upon binding to dIgA is another open issue. Moreover, structural differences between human IgA1 and IgA2 exists (Boehm *et al.*, 1999; Furtado *et al.*, 2004) which are consistent with findings showing distinct binding modes of human SC to dIgA1 and dIgA2 (Almogren *et al.*, 2007a). Knowledge of the three-dimensional structure of SC represents a mandatory first step toward the resolution of the complete structure of SIgA. If we assume that the model of SC is not that different to the extracellular portion of pIgR, it might also be helpful *per se* to map binding site(s) to permit epithelial drug targeting through specific interaction with pIgR itself (Ferkol *et al.*, 1995; White & Capra, 2002; Braathen *et al.*, 2006).

4.4 Materials and Methods

4.4.1 Preparation of recombinant SC and its fragments

Recombinant human SC was produced in transfected Chinese hamster ovary cells and purified as reported (Phalipon *et al.*, 2002). The SC samples were kindly prepared by Dr. Clementine Perrier and Dr Blaise Corthésy. Filtered aliquots were stored at 4°C in PBS (116 mM NaCl, 10.4 mM Na_2HPO_4, 3.2 mM NaH_2PO_4, pH 7.2) prior to use for scattering and ultracentrifugation. Cleaved SC fragments were obtained from purified SC samples that were stored for at least 6 months at 4°C. Separation was performed by chromatography onto a 1 m x 1.6 cm Superdex-200 column (Amersham-Pharmacia) in PBS. Pools of fractions containing fragments D1-D3 and D4-D5 were concentrated using Amicon Ultra (30 kDa molecular weight cut-off) and Centricon YM-10 (10 kDa molecular weight cut-off) filters respectively (Millipore). Further purification of fragments D1-D3 and D4-D5 was achieved by repeating chromatography onto the Superdex-200 column. To identify the SC cleavage site, N-terminal sequencing of purified D4-D5 from a polyvinylidene difluoride blot was performed by Edman degradation at Alta Biosciences (University of Birmingham, UK).

For the SAXS and AUC experiments, samples were recovered in PBS. For neutron scattering, the buffer was Dulbecco's PBS (137 mM NaCl, 2.7 mM KCl, 8.1 mM Na_2HPO_4, 1.5 mM KH_2PO_4, pH 7.5), made up in 100% ^2H_2O. The neutron samples were dialysed at 6°C into this 100% ^2H_2O buffer for 36 h with four buffer changes. The two buffer densities in H_2O solvent were predicted to be 1.00485 g/ml and 1.00543 g/ml respectively from these compositions using the program SEDNTERP (Laue *et al.*, 1992). The SC, D1-D3 and D4-D5 concentrations were determined using absorption coefficients of 12.3, 7.1 and 18.7 (1%, 280 nm, 1 cm path length) that were calculated from its composition by assuming the presence of seven biantennary complex-type oligosaccharides Man3GlcNAc4Gal2NeuNAc2 (Perkins, 1986; Figure 4.1). The partial specific volume $\bar{v}$ for SC, D1-D3 and D4-D5 was calculated to be 0.717 ml/g, 0.717 ml/g and 0.718 ml/g respectively from their compositions (Perkins, 1986; Section 3.3.1). The calculated molecular mass values for SC, D1-D3 and D4-D5 were 79.6 kDa, 45.5 kDa and 34.1 kDa respectively, their unhydrated volumes were 99.7 nm^3, 56.9 nm^3 and 42.7 nm^3

respectively, and their hydrated volumes were 132.2 nm^3, 75.5 nm^3 and 56.7 nm^3 respectively (assuming 0.3 g H_2O/g glycoprotein).

4.4.2 X-ray and neutron scattering data collection

SAXS data was obtained during two independent sessions on the Beamline ID02 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France at a ring energy of 6.0 GeV (Narayanan *et al.*, 2001; Section 2.2.3.2). The experiments were performed in 16-bunch mode with a storage ring current ranging from 92 mA to 53 mA (SC) and 4-bunch mode with a storage ring current ranging from 41 mA to 23 mA (D1-D3 and D4-D5). The final data analyses were performed with a sample-to-detector distance of 3.0 m, which yielded a Q range from 0.07 nm^{-1} to 2.1 nm^{-1}. Protein samples were measured at three or four concentrations between 0.62 mg/ml to 2.0 mg/ml (SC), 0.39 mg/ml to 1.6 mg/ml (D1-D3) and 0.40 mg/ml to 0.73 mg/ml (D4-D5). Data were acquired using five or ten time-frames, with each frame lasting for 0.5, 1, 2 or 4 seconds. Comparisons of the first time frame with the fully-merged ten time-frames were used to establish the absence of radiation damage. Data reduction involved the calibration of the detector intensities using a Lupolen sample and the Q range using a silver behenate salt standard respectively.

Neutron scattering data for SC was collected in one session on Instrument LOQ at the pulsed neutron source ISIS at the Rutherford Appleton laboratory, Didcot, UK (Heenan *et al.*, 1997; Section 2.2.3.3). The pulsed neutron beam was derived from a proton beam current of 172-173 μA. Data acquisition were for fixed monitor counts of 80×10^6 or 160×10^6, during runs which lasted 1-2 hours for concentrations between 0.62 mg/ml to 2.0 mg/ml, measured at 15 °C. Other details including the Guinier analyses to determine the radius of gyration R_G, the cross-sectional radius of gyration R_{XS} and the calculation of the distance distribution function $P(r)$ using GNOM, for which the Q range was 0.08-2.06 nm^{-1} are described in Section 2.2.4.

4.4.3 Sedimentation equilibrium and sedimentation velocity data acquisition

AUC was performed at 20°C on a Beckman XL-I instrument, equipped with an AnTi50 rotor. SE data sets for SC were acquired over 45 h using six-sector cells with solution column heights of 2 mm at rotor speeds of 8,000 r.p.m., 11,000 r.p.m., 14,000 r.p.m., 18,000 r.p.m. and 23,000 r.p.m, until equilibrium had been reached at each speed as shown by the perfect overlay of runs measured at five hour intervals (Section 2.3.4). Nine concentrations of SC between 0.20 mg/ml to 2.50 mg/ml were studied using absorbance optics at 280 nm and by interference optics. SV data for SC, D1-D3 and D4-D5 was acquired over 16 h at rotor speeds of 20,000 r.p.m., 25,000 r.p.m., 30,000 r.p.m, 35,000 r.p.m (and 42,000 r.p.m. for D1-D3 and D4-D5) in two-sector cells with column heights of 12 mm (Section 2.3.3). SC was studied at five concentrations from 0.30 mg/ml to 2.50 mg/ml. D1-D3 and D4-D5 were each studied at two concentrations of 0.39 mg/ml to 1.6 mg/ml and 0.40 mg/ml to 0.73 mg/ml respectively. For the DCDT+ *g(s*)* time-derivative analyses for SC and D4-D5, successive absorbance and interference scans were recorded at 5 min intervals at 280 nm, the shortest interval possible under standard measurement conditions. The optimal rotor speeds were 30,000 r.p.m. to 42,000 r.p.m. for the interference and absorbance data sets. The subtraction of pairs of boundary scans *versus* cell radius eliminates systematic errors from baseline distortions in the cell windows and permits the averaging of several pairs of subtractions. The extrapolation of individual subtractions to the start time gives the *g(s*)* function, which was computed using DCDT+ version 1.16 (Philo, 2000). The sedimentation and diffusion coefficients were determined using between 4 to 20 scans were used in order to ensure that the time broadening effects were negligible. The alternative approach to determine $s^{\circ}_{20,w}$ is to perform direct boundary fits of up to 700 absorbance and interference scans to the Lamm equation using SEDFIT (Schuck, 1998). In this, *c(s)* size-distribution analyses were computed in order to identify the species that were sedimenting in the sample, from which the $s^{0}_{20,w}$ values were determined from the observed peak positions. The final SEDFIT analyses were optimised until the r.m.s.d and visual appearance of the fits and the residual bitmaps were satisfactory.

4.4.4 Conformational searches for the best-fit solution model of SC

The SC sequence was taken from residues 19-603 of the polymeric immunoglobulin receptor (SWISSPROT code P01833) (Figure 4.1b). Using this sequence, BLASTP was employed to identify a known structural template with the closest sequence similarity to SC. Consequently, the homology modelling of the D2-D5 domains was based on a topological sequence alignment with the crystal structure of the D1 domain at 0.19 nm resolution (PDB code 1XED; Hamburger *et al.*, 2004; Section 3.2.1). The secondary structure of the crystal structure was identified using DSSP (Kabsch & Sander, 1983). The start and end of domains D2 to D5 is defined as residues 1 to 104 in Figure 4.1. Homology models of domains D2-D5 were created using MODELLERv7 (Sali & Blundell, 1993). The seven *N*-linked biantennary oligosaccharide chains $Man_3GlcNAc_4Gal_2NeuNAc_2$ were incorporated into these D1 to D5 structures using procedures described for IgA1 and IgA2 and based on the oligosaccharide structure seen in human IgG1 Fc (Boehm *et al.*, 1999; Furtado *et al.*, 2004; Deisenhofer, 1981).

In order to model the interdomain linker structures, the C-terminal peptide 543-AVEER....AEEKA-585 and the four interdomain linker peptides 109-VSQGPGLL-116, 219-VLK-221, 324-WQLFVNEESTIR-335 and 440-IEGEPNL-446 were initially created in an extended β-strand conformation with an additional residue at the start and end as required. INSIGHT II 98.0 molecular graphics software with BIOPOLYMER, DISCOVER, HOMOLOGY and DELPHI modules (Accelrys, San Diego, CA, USA) was used on Silicon Graphics O2 and OCTANE Workstations. A total of 5,000 random linker conformations for each of the five peptides were generated from the starting β-strand model by a molecular dynamics simulation using DISCOVER3 (Boehm *et al.*, 1999; Aslam *et al.*, 2003). A wide range of conformations was generated by performing the simulation at a temperature of 773 K. First, the peptide structure was subjected to energy minimisation for 300 iterations. After a temperature equilibration step of 5,000 fs, the simulation was run for 500,000 fs. The linker structure was saved every 100 fs to produce the linker library. Length constraints were applied when required in order to limit the conformational variability of the linkers. As residues SP of the N-terminal peptide KSP were already observed in the D1 crystal structure, the first residue K was added to the crystal structure using MODELLER v7 (Sali & Blundell, 1993).

Each full model for SC was created in Search 1 by superimposing the four main-chain atoms of each N-terminal and C-terminal linker residue onto the D1-D5 domain in question. After this, the duplicated residues were deleted from the SC model prior to the scattering calculations. Models for D1-D3 and D4-D5 were created independently from those for SC in up to five conformational searches. The final D1-D3 and D4-D5 models were then used in Search 2 for SC models in which only the 324-WQLFVNEESTIR-335 linker was varied.

4.4.5 Debye scattering and sedimentation coefficient modelling of SC

Each SC, D1-D3 and D4-D5 model was converted into Debye spheres in order to calculate X-ray and neutron scattering curves for comparison with the experimental curves (Section 3.3.3). For SC, D1-D3 and D4-D5, trial runs showed that a cube side length of 0.542 nm in combination with a cutoff of 4 atoms consistently produced sphere models within 95% of the total unhydrated volume of 99.7 nm^3, 56.9 nm^3 and 43.8 nm^3 respectively calculated from their compositions. Hence the optimal totals of unhydrated spheres for the neutron modelling of SC, D1-D3 and D4-D5 were 626, 268 and 353 respectively. Since the hydration shell surrounding glycoproteins is detected by X-ray scattering, the sphere models were accordingly adapted by adding spheres to the surface of the unhydrated models using HYPRO (Ashton *et al.*, 1997) based on 0.3 g H_2O / g glycoprotein to result in a total hydrated volume of 132.2 nm^3 for SC, 75.5 nm^3 for D1-D3 and 56.7 nm^3 for D4-D5. The optimal totals of hydrated spheres for SC, D1-D3 and D4-D5 were 830, 474 and 356 respectively. The calculations for the generation of the Debye scattering curves *I(Q)* for the models is give in Section 3.3.4. A flat background correction of 2.7% of *I(0)* was applied to the final neutron scattering curve fits in Figure 4.13 and Figure 4.14 to allow for a uniform incoherent scattering contribution from residual protons in the sample (Section 2.2.2).

The number of spheres N in the unhydrated and hydrated models after grid transformation was used to assess steric overlap between the D1 – D5 domains, where models showing less than 95% of the optimal total were discarded. The models were initially assessed by calculation of the X-ray and neutron R_G and R_{XS} values in the same Q ranges used for the experimental Guinier fits. Models that passed these R_G and R_{XS} filters were then ranked using a goodness-of-fit agreement

with the experimental curves (*R*-factor), defined by analogy with protein crystallography and based on the *Q* range extending to 2.0 nm^{-1}. (Beavil *et al.*, 1995)

The sedimentation coefficients $s^{o}_{20,w}$ were calculated directly from the SC, D1-D3 and D4-D5 PDB files using HYDROPRO. The hydration shell surrounding SC and its two fragments was represented by the recommended value of 0.31 nm for the atomic element radius for all atoms as an empirical approximation (Sun *et al.*, 2004; Garcia de la Torre *et al.*, 2000)

4.4.6 SC Protein Data Bank accession code

The ten best-fit SC α-carbon co-ordinate models have been deposited in the PDB with the accession codes 2OCW (Figure 4.16; Table 4.2).

Chapter Five

Implications of the Near-Planar Solution Structure of Human Myeloma Dimeric IgA1 for Mucosal Immunity and IgA Nephropathy

5.1 Introduction

Immunoglobulin A (IgA) is the most abundant antibody class found in mucosal surfaces and is also the most heterogeneous human antibody class. IgA occurs in serum predominantly as monomeric IgA1 (with around 13% polymeric IgA) and at mucosal surfaces as (usually) dimeric IgA (dIgA) covalently linked with secretory component, with generally slightly more IgA1 than IgA2 (Kerr, 1990). dIgA interacts with the cell-surface Fcα receptor (FcαRI; CD89), which is present on the surface of neutrophils, monocytes, macrophages and eosinophils to perform immune functions (Morton *et al.*, 1996). An increased polymeric IgA1 level in serum and mesangial polymeric IgA1 depositions can be associated with IgA nephropathy (IgAN) which is the most common form of chronic glomerulonephritis world-wide (Barratt & Feehally, 2005).

This chapter focuses on the structure of myeloma dIgA1 in serum (Figure 5.1). The three-dimensional structures of the different forms of polymeric IgA are not well understood. Structures for intact IgA at medium resolution for monomeric human IgA1 and IgA2 together with the covalent adduct of IgA1 with human serum albumin have been determined by constrained scattering modelling (Boehm *et al.*, 1999; Furtado *et al.*, 2004; Almogren *et al.*, 2006). The overall structure of human myeloma dIgA has been observed by electron microscopy as a tail-to-tail arrangement of the Fc regions in the two IgA monomers (Bloth & Svehag 1971; Munn *et al.*, 1971; Feinstein *et al.*, 1971; Nezlin, 1998; Figure 1.9). However the conditions of electron microscopy work using heavy metal stains *in vacuo* can be harsh, and the solution arrangement of the monomers in dIgA1 is unknown up to now. The size of dIgA1, together with its high glycosylation and long flexible hinge and tailpiece regions, suggest that its crystal structure will be hard to achieve (Figure 5.1; Figure 5.2). In such a situation, X-ray and neutron scattering and analytical ultracentrifugation combined with constrained modelling, will result in a medium resolution structure determination that will reveal the first molecular details of the two IgA monomers within dIgA (Perkins *et al.*, 2008). Myeloma dIgA1 is shown to be formed as a near-planar molecule in solution with a bent arrangement of Fc regions in which the J chain does not act as a spacer between them. Its structure rationalises the accommodation of the J chain within dIgA1, as well as several known effector functions of dIgA1, particularly its interaction with FcαRI receptors on cells of the immune

system, and the interaction with secretory component in relation to the formation of SIgA. Insight is also provided into the role of dIgA1 in the development of IgAN.

5.2 Results

5.2.1 Purification of human myeloma dIgA1

The conventional purification of polymeric IgA from normal human serum is lengthy and labour intensive, and is inadequate for purifying the polymeric forms of IgA. (Delacroix *et al.*, 1982; Loomes *et al.*, 1991). A procedure based on thiophilic resins with a binding capacity of over 28 mg of immunoglobulins per ml of resin allows rapid purification of immunoglobulins which bind selectively to the resin whilst almost all the other serum proteins do not (Sun *et al.*, 2005; Almogren *et al.*, 2007).

For the work presented in this chapter, dIgA1 was purified from anonymised human myeloma serum selected for its high polymer content (Figure 5.3). The myeloma serum had an immunoglobulin content of IgA at 15 g/l, with a dimer to monomer ratio of 2:1, while IgG and IgM were present at 7 g/l and 1.5 g/l respectively. dIgA1 was shown by non-reducing SDS-PAGE to be well resolved from monomeric IgA1 in subsequent S-300 gel filtration (Figure 5.3a,b). Reducing SDS-PAGE showed the expected bands corresponding to the IgA1 heavy chain at 62 kDa, and the light chain at 30 kDa (Figure 5.3c). Monoclonal anti-κ light chain and anti-human J chain antibodies were used in Western blots to confirm the presence of dIgA1 (Figure 5.3d). In Sephacryl S300 size-exclusion gel filtration, dIgA1 eluted as a single symmetrical peak (data not shown). The dIgA1 sample was concentrated for scattering and ultracentrifugation data collection.

5.2.2 X-ray and neutron scattering analyses of dIgA1

The averaged solution arrangement of the two IgA1 monomers within dIgA1 was analysed by comparing X-ray scattering data *I(Q)* for dIgA1 with previous X-ray data for serum IgA (Boehm *et al.*, 1999; Furtado *et al.*, 2004; Almogren *et al.*, 2006). Data were collected at dIgA1 concentrations between 0.6 mg/ml to 1.2 mg/ml at the ESRF. The comparison of data from a single time frame exposure with the average over ten consecutive time frames showed that the X-ray Guinier analyses were unaffected by

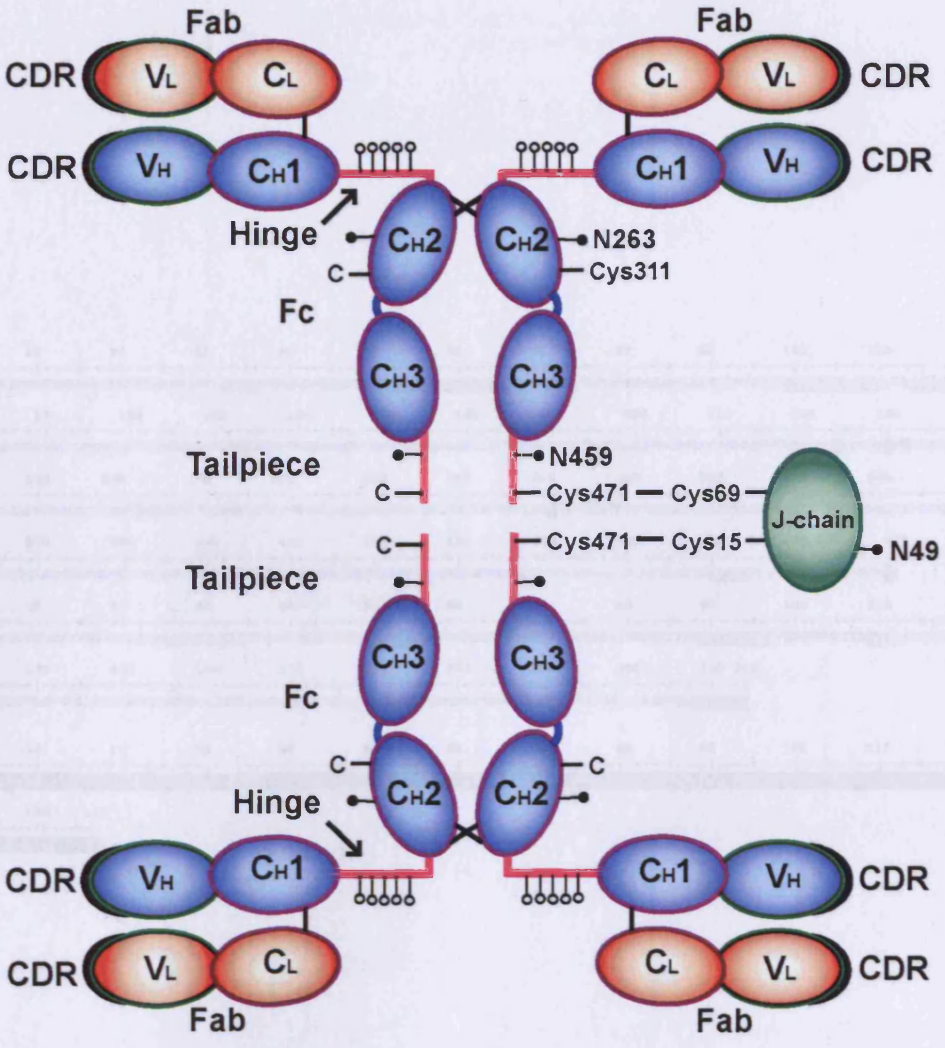

Figure 5.1 The domain structure of dIgA1

dIgA1 is formed by two IgA1 monomers that are covalently linked with the J-chain (green). Each IgA1 heavy chain (blue) contains the V_H, C_H1, C_H2 and C_H3 domains, and each light chain (red) contains the V_L and C_L domains. The variable (V) and constant (C) domains are highlighted with green and purple borders respectively. The complementarity-determining regions (CDR: black crescents), the 23-residue hinge peptide (pink), and the 18-residue C-terminal tailpiece (blue) are highlighted. Inter-domain disulphide bridges are shown as an X. Cys471 in one tailpiece of each Fc region is disulphide-bridged with Cys15 and Cys69 in the J chain. The N-linked oligosaccharide sites (●) at Asn263 and Asn459 on dIgA1 and Asn49 on J chain, and the five O-linked oligosaccharides in each hinge (○) are shown.

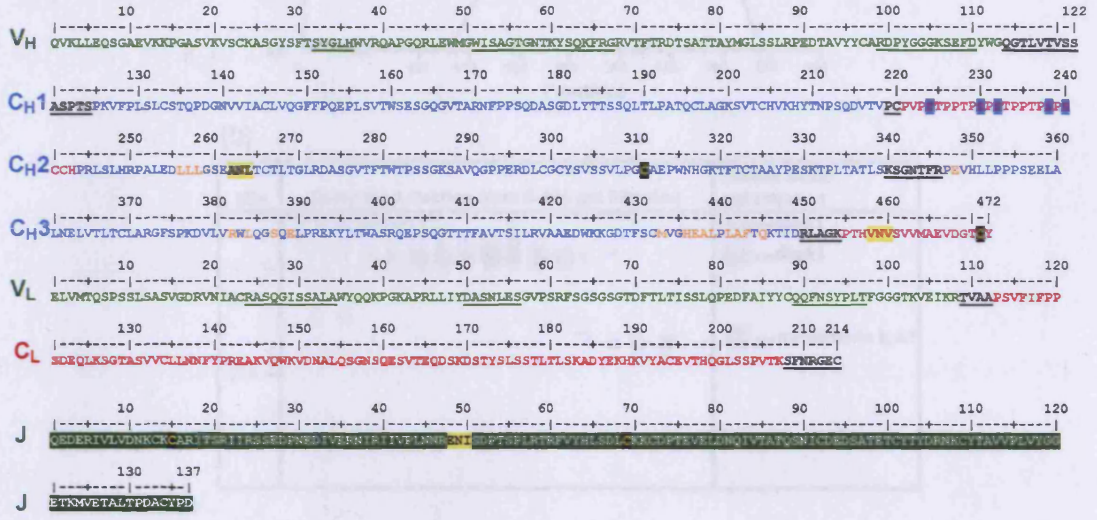

Figure 5.2 The sequence of an IgA1 monomer and J chain

The IgA1 variable domains (green), heavy chain constant domains (blue) and light chain constant domains (red) are separated by linkers (black underlined). The hinge region (pink) and tailpiece (pink), the six sets of CDR residues (underlined), the FcαRI receptor binding residues (orange), two N-linked oligosaccharide sites (yellow background) and five O-linked oligosaccharide sites (blue background), and Cys311/Cys471 (yellow with black background) are highlighted. The residue numbering of the IgA heavy chain constant domains follows that of IgA1 Bur. The J chain sequence (white with green background) is shown, with the N-linked oligosaccharide (black with yellow background) and Cys15/Cys69 (yellow with black background) is highlighted.

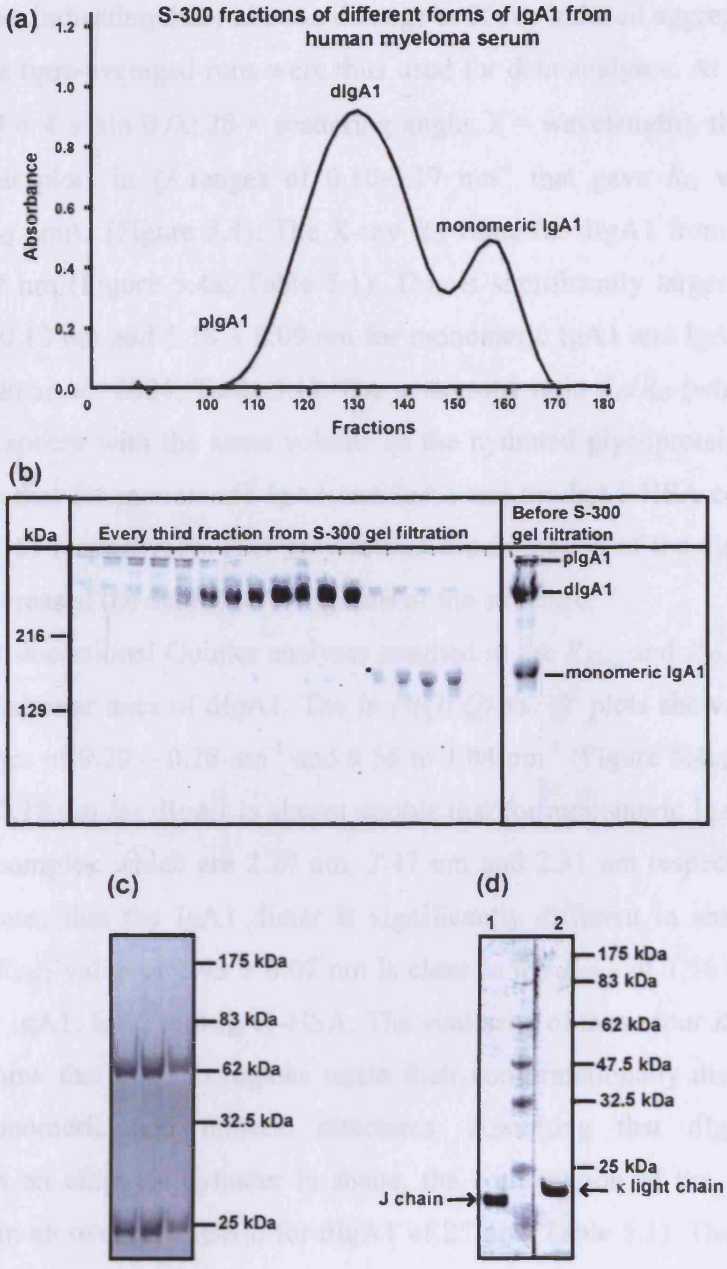

Figure 5.3 Purification of human myeloma dIgA1

(a) Chromatogram of an S-300 gel filtration column showing separation of the different molecular forms of IgA1 from the serum of a myeloma patient. (b) Western blot using an anti-IgA1 Ab of non-reducing SDS-PAGE for every third S-300 eluted fraction from (a). (c) Reducing SDS-PAGE of the dIgA1 peak fractions. (d) Western blot of the dIgA1 preparation after S-300 gel filtration using (1) anti-J chain or (2) anti-κ lght chain Abs. (Taken from Almogren, 2003).

the exposure time, indicating that radiation damage or X-ray induced aggregation effects were absent. The time-averaged runs were thus used for data analyses. At the lowest Q values (where $Q = 4\ \pi \sin\theta / \lambda$; 2θ = scattering angle; λ = wavelength), these analyses resulted in linear plots in Q ranges of 0.10-0.17 nm^{-1} that gave R_G values within satisfactory $Q.R_G$ limits (Figure 5.4). The X-ray R_G value for dIgA1 from eight values was 8.65 ± 0.27 nm (Figure 5.4a; Table 5.1). This is significantly larger than the R_G value of 6.20 ± 0.13 nm and 5.18 ± 0.09 nm for monomeric IgA1 and IgA2 (Boehm *et al.*, 1999; Furtado *et al.*, 2004; Table 5.1). The anisotropy ratio R_G/R_O (where R_O is the R_G value of the sphere with the same volume as the hydrated glycoprotein) for dIgA1 was 2.16, where that for monomeric IgA1 and IgA2 and the IgA1-HSA complex were 1.99, 1.66 and 2.13 respectively. This showed that the formation of the dimer from the monomer had increased the degree of elongation of the structure.

X-ray cross-sectional Guinier analyses resulted in the $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ parameters that monitor the shorter axes of dIgA1. The $ln\ (I(Q).Q)$ vs. Q^2 plots showed two linear region in Q ranges of 0.20 – 0.28 nm^{-1} and 0.56 to 1.04 nm^{-1} (Figure 5.4c,d). The $R_{XS\text{-}1}$ value of 3.94 ± 0.18 nm for dIgA1 is almost double that for monomeric IgA1, IgA2 and the IgA1-HSA complex which are 2.20 nm, 2.47 nm and 2.31 nm respectively (Table 5.1). This indicates that the IgA1 dimer is significantly different in shape from the monomer. The $R_{XS\text{-}2}$ value of 1.43 ± 0.07 nm is close to the $R_{XS\text{-}2}$ of 1.56 nm, 1.47 nm and 1.31 nm for IgA1, IgA2 and IgA1-HSA. The similarity of these four $R_{XS\text{-}2}$ values is interpreted to show that the Fab regions retain their conformationally distinct identity within the monomeric and dimeric structures. Assuming that dIgA1 can be approximated as an elliptical cylinder in shape, the combination of the R_G and $R_{XS\text{-}1}$ values resulted in an overall length L for dIgA1 of 25 nm (Table 5.1). The comparison of this length with that of 20.1 nm for IgA1 and 16.1 nm for IgA2 showed that the increased size of dIgA1 does not correspond to a simple extended end-to-end association of the longest axes of the monomers.

Neutron scattering data were obtained for dIgA1 at 0.6-1.5 mg/ml in 100% ^2H_2O buffer as a control for possible X-ray-induced radiation damage. This buffer corresponds to a high negative protein-solvent scattering contrast, instead of the high positive contrast observed by X-rays, hence acts as a control of any internal scattering

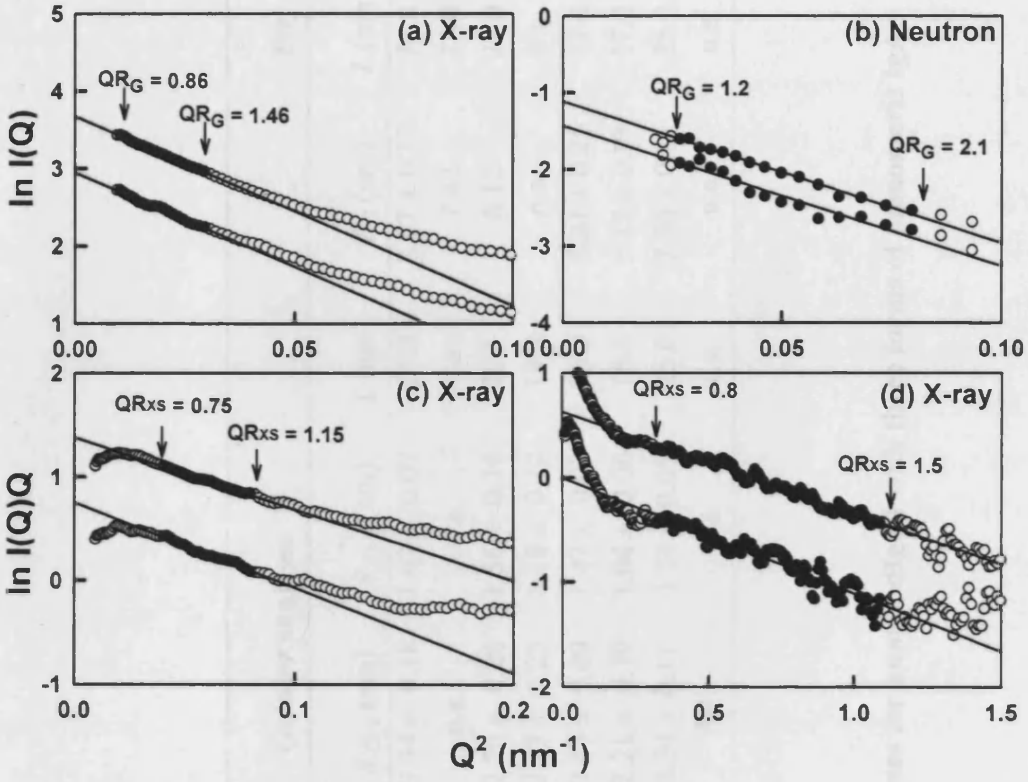

Figure 5.4 X-ray and neutron Guinier analyses for dIgA1

The $Q.R_G$ and $Q.R_{XS}$ ranges used to determine the R_G, $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ values are represented by filled circles between the arrowed data points. (a,b) The X-ray and neutron Q ranges used for the R_G analyses were 0.10 to 0.17 nm^{-1}. (c,d) The Q ranges used to calculate the $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ values were 0.20 to 0.28 nm^{-1} and 0.56 to 1.04 nm^{-1} respectively. In (a, c, d), the X-ray fits for dIgA1 correspond to concentrations of 1.16 mg/ml (upper) and 0.58 mg/ml (lower. In (b), the neutron fits for dIgA1 correspond to 1.16 mg/ml (upper) and 0.83 mg/ml (lower).

Protein	Data	Guinier analyses				*P(r)* analyses			
		R_G (nm)	R_{XS-1} (nm)	R_{XS-2} (nm)	L (nm)	R_G (nm)	L (nm)	*M1* (nm)	*M2* (nm)
Dimeric IgA1	X-ray	8.65 ± 0.27	3.94 ± 0.18	1.43 ± 0.07	25.0	8.67 ± 0.17	26.0	4.9	9.9
	Neutrons	7.60 ± 0.05	n.a.	n.a.	n.a.	7.42	23.0	5.1	10.1
Monomeric IgA1 (Boehm *et al.*, 1999)	X-ray	6.20 ± 0.13	2.20 ± 0.26	1.56 ± 0.16	20.1	6.12	21.0	3.7	9.1
	Neutrons	6.11 ± 0.18	2.17 ± 0.23	1.18 ± 0.12	19.8	n.a.	n.a.	n.a.	n.a.
Monomeric IgA2m(1) (Furtado *et al.*, 2004)	X-ray	5.18 ± 0.09	2.47 ± 0.09	1.47 ± 0.08	16.0	5.21 ± 0.15	17.0	4.5	6.8
	Neutrons	5.03 ± 0.01	2.21 ± 0.10	1.04 ± 0.06	15.5	5.12 ± 0.01	17.0	4.5	n.a.
IgA1-HSA (Almogren *et al.*, 2006)	X-ray	7.47 ± 0.31	2.31 ± 0.11	1.31 ± 0.05	25.0	7.36 ± 0.01	25.0	4.9	8.3
	Neutrons	7.09 ± 0.01	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

Table 5.1 Comparison of Guinier and *P(r)* analyses for human dIgA1 with three forms of monomeric IgA

n.a. – not available

inhomogeneity effects caused by the estimated 9.6% carbohydrate content of dIgA1. Neutron scattering data also provided a different view of dIgA1 in that the hydration shell surrounding dIgA1 is visible by X-ray scattering, but not by neutron scattering (Perkins, 2001). The dIgA1 neutron R_G from two values was 7.60 ± 0.05 nm (Figure 5.4b; Table 5.1). This value supports the X-ray R_G determination. However this is 1 nm less than the X-ray R_G value, and the difference is larger than seen previously with IgA1 and IgA2 (Boehm *et al.*, 1999; Furtado *et al.*, 2004). The non-visibility of the hydration shell when visualised by neutrons has a large effect, and this is attributed to the positioning of most of the hydration shell at large distances from the centre of the dIgA1 structure. The neutron anisotropy ratio R_G/R_O was 2.08, and is similar to the X-ray-derived value to confirm an elongated conformation of dIgA1 in solution. As the consequence of poor counting statistics at large Q, no neutron $R_{XS\text{-}1}$ or $R_{XS\text{-}2}$ values could be determined for dIgA1. There is a linear relationship between the LOQ Guinier $I(0)/c$ values for proteins measured in ^2H_2O buffers and normalised against a standard deuterated polymer and the protein molecular mass M_r, where $M_r = I(0)/c \times 9.10^5$ (Boehm *et al.*, 1999). The dIgA1 neutron Guinier $I(0)/c$ value of 0.33 ± 0.05 resulted in a M_r value of 297 ± 45 kDa (Figure 2.5). This measurement is within error of the composition-derived value of 345 kDa for dIgA1, showing that the scattering data corresponded well to the expected composition of dIgA1 with nine N-linked and 20 O-linked biantennary oligosaccharides (Figure 5.1; Figure 5.2).

Structural dimensions are provided by the transformation of the $I(Q)$ curve into the distance distribution function $P(r)$ curve. The mean R_G values determined from the X-ray and neutron $P(r)$ curves were 8.67 ± 0.17 nm (6 values) and 7.47 nm (2 values) respectively. These R_G values were therefore consistent with the Guinier R_G values (Table 5.1). The X-ray $P(r)$ curve for dIgA1 was of better quality than the neutron $P(r)$ curve for reason of better signal-noise ratios (Figure 5.5). Nonetheless the two curves were reproducible in that both showed a double peak with maxima of *M1* and *M2* that were each located at similar r values of 4.9-5.1 nm and 9.9-10.1 nm respectively (Table 5.1). These two peaks correspond to an abundance of interatomic vectors within the dIgA1 solution structure. The dIgA1 peak *M1* at 4.9-5.1 nm was also evident in the $P(r)$ curve for monomeric IgA1 and IgA2 and for myeloma IgA1-HSA at 3.7 nm, 4.5 nm and

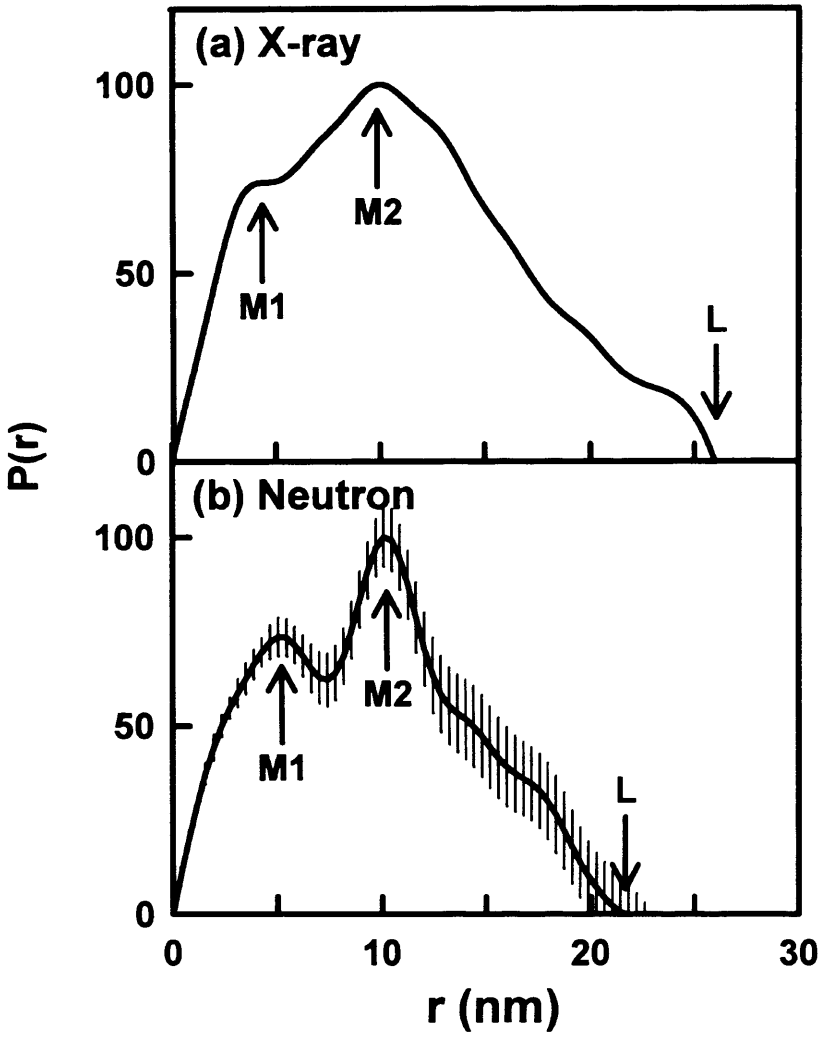

Figure 5.5 X-ray and neutron distance distribution functions *P(r)* for dIgA1

The concentrations of dIgA1 were 1.15 mg/ml (X-rays) and 1.16 mg/ml (neutrons). The maxima *M*1 and *M*2 of the *P(r)* curve depict two frequently occurring distance within the dIgA1 structure. The positions of *M*1 and *M*2 are 4.94 nm and 9.88 nm respectively (X-rays) and 5.06 and 10.14 nm respectively (neutrons). The length of IgA1 is denoted by *L* at the *r* value where the *P(r)* curve reaches 0. The *L* values for dIgA1 were 26 nm (X-rays) and 23 nm (neutrons).

4.9 nm respectively (Table 5.1). This *M1* peak is assigned to the most commonly occurring distance within a single Fab or Fc region, given that each is approximately 8 nm long. Hence the *M1* peak in Figure 5.5 is consistent with the well-defined existence of the Fab region within the solution structure of dIgA1. The peak *M2* at 9.9-10.1 nm is of greater intensity than the peak *M1* when this is compared with monomeric IgA1 and IgA2 (Boehm *et al.*, 1999; Furtado *et al.*, 2004). The reason for this intensity difference may be that there are a large number of interatomic vectors between the four Fab regions in the dimer that are close to 10 nm and held in a comparatively inflexible arrangement. The maximum length L is determined from where the $P(r)$ curve reaches zero at large r values, and L is determined independently of shape assumptions. The dIgA1 X-ray and neutron maximum length L was found to be 26 nm and 23 nm respectively. The shorter L value with neutrons indicates that a significant proportion of the hydration shell is located at the periphery of the dIgA1 structure. Simple geometric considerations of planar end-to-end arrangements of two monomer structures suggested by Figure 5.1 readily leads to L values slightly greater than 26 nm. Hence it can be concluded that the dIgA1 solution structure is composed of an extended arrangement of two monomers joined at the end of their Fc regions, and that the two monomers are not positioned as a more compact solution structure that might resemble a bouquet of four Fab flowers positioned on two Fc stalks.

5.2.3 Analytical ultracentrifugation of dIgA1

In order to confirm the molecular mass and oligomerisation state of dIgA1, sedimentation equilibrium (SE) experiments were performed at six concentrations between 0.07 mg/ml and 1.0 mg/ml and at six rotor speeds using both interference optics and absorbance optics at 280 nm. On the assumption that a single species was present, SE fits were carried out with nine curves at each dIgA1 concentration using both data types (Section 5.4.3). The individual fits gave molecular mass values that ranged between 280 kDa to 345 kDa (Figure 5.6a,b). Their extrapolation to zero concentration gave a mean molecular mass of 335 kDa (± 15) kDa (Figure 5.6c). These are within error of the molecular mass of 345 kDa calculated from the sequence of dIgA1 if the 29 N-glycosylation sites in this are all occupied by biantennary oligosaccharides. Had the

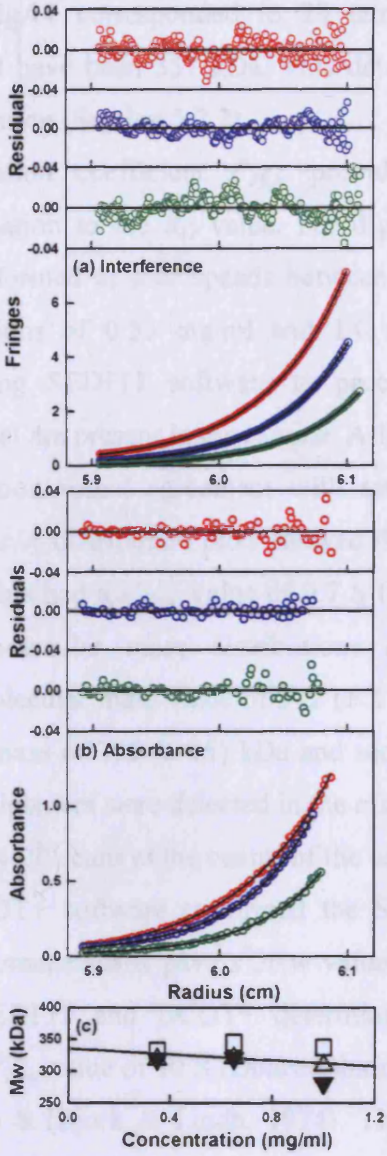

Figure 5.6 Sedimentation equilibrium analyses for dIgA1

dIgA1 was analysed using (a) interference and (b) absorbance optics at concentrations of 1.0 mg/ml (red), 0.65 mg/ml (blue) and 0.35 mg/ml (green). Circles represent the experimental data and the continuous lines represent their fits. The curve fit residuals are shown in the panels above the exponential fits, using the same colour scheme. The (c) panel below shows any dIgA1 concentration and speed dependence of the observed molecular weights. The rotor speeds are denoted by (□) 8,000 r.p.m., (△) 11,000 r.p.m., and (▼) 14,000 r.p.m. (c) Calculations of the best-fit line and its extrapolation to zero gave a dIgA1 molecular mass value of 335.0 ± 15 kDa.

N-glycosylation of dIgA1 corresponded to 29 tetraantennary oligosaccharides, the molecular mass would have been 357 kDa. This determination confirmed the neutron molecular mass calculation (Section 5.2.2).

The sedimentation coefficient $s^o_{20,w}$ provides an independent measure of macromolecular elongation to the R_G value. For dIgA1, sedimentation velocity (SV) experiments were performed at four speeds between 10,000 r.p.m. and 25,000 r.p.m. using two concentrations of 0.53 mg/ml and 1.0 mg/ml (Section 5.4.3). The *c(s)* distribution plots using SEDFIT software to process all the scans identifies the sedimenting species that are present in the sample. All the experimental and fitted dIgA1 boundaries showed good visual agreement with satisfactory r.m.s.d. values (Figure 5.7a,b). The resulting *c(s)* distribution plots showed that only a single species of dIgA1 was present and that this had a $s^o_{20,w}$ value of 9.7 S (Figure 5.7c,d). The conversion of the *c(s)* plots to molecular mass distributions *c(M)* showed that this species correspondeds to a molecular mass value of 342 (± 11) kDa. This agreed well with the SE dIgA1 molecular mass of 335 (± 15) kDa and sequence-derived molecular mass of 345 kDa. No higher oligomers were detected in the *c(s)* plots (Figure 5.7c,d). A different SV analysis based on 4-20 scans at the centre of the sedimentation experiment using the *g(s*)* method in DCDT+ software confirmed the SEDFIT results. Here, the dIgA1 absorbance and interference scans gave $s^0 20,w$ values of 9.8 S and 9.9 S respectively (Figure 5.8). The SEDFIT and DCDT+ determinations agree well with previous literature reporting a $s^o_{20,w}$ value of 10 S (Dourmashkin *et al.*, 1971), 9.65 S (Feinstein *et al.*, 1971) and 9.2-9.3 S (Bjork & Lindh, 1974). The sedimentation coefficient $s^o_{20,w}$ leads to the frictional ratio f/f_O, where f_O is the f value of the sphere with the same volume as the hydrated glycoprotein. Its value for dIgA1 of 1.63 is slightly larger than those of 1.56 found for monomeric IgA1 and 1.53 for monomeric IgA2 (Furtado *et al.*, 2004). This indicates that dIgA1 is more elongated than IgA1. This is consistent with the X-ray and neutron scattering results reported above, and shows that dIgA1 is formed as an extended arrangement of two IgA1 monomers.

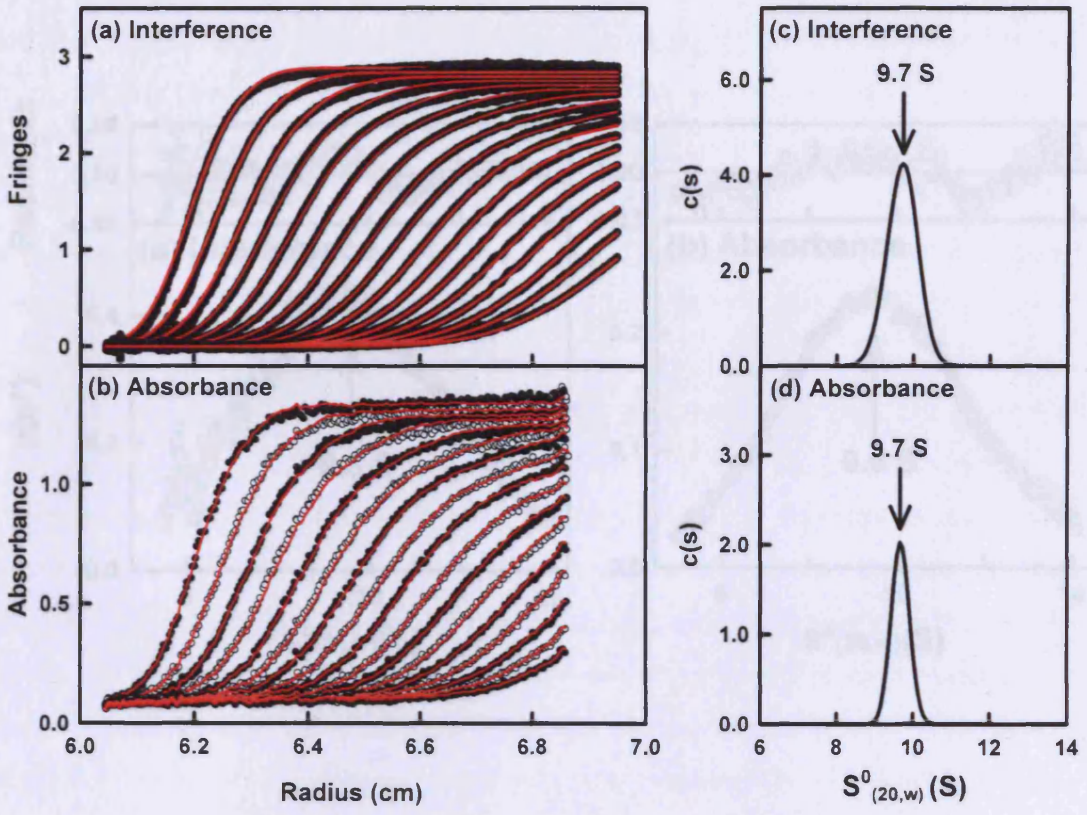

Figure 5.7 dIgA1 sedimentation velocity *c(s)* distribution analyses using SEDFIT
(a,b) The circles represent the experimental data and the continuous red lines represent the fits, where only every sixth scan of the 120 scans used in the Lamm fits is shown for reason of clarity. The interference scans for dIgA1 at a concentration of 0.96 mg/ml and a rotor speed of 20,000 r.p.m. are shown in (a), while the corresponding absorbance scans at 280 nm are shown in (b). The scans were recorded at 5 min intervals. (c,d) The corresponding *c(s)* plots are shown from which the sedimentation coefficient for dIgA1 was determined to be 9.7 S.

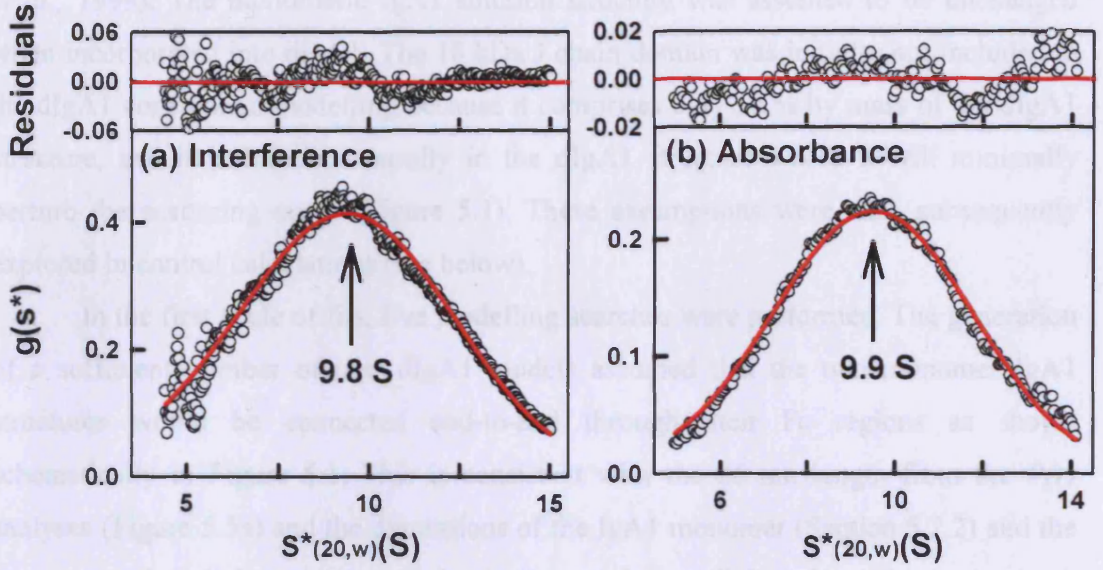

Figure 5.8 Time-derivative sedimentation velocity analyses of dIgA1 using DCDT+
In the *g(s*)* plots, the experimental data are represented by circles, and the fits are shown by the continuous red lines. The goodness-of-fit residuals are shown above each *g(s*)* analysis. The arrowed $s_{20,w}$ values for dIgA1 using (a) interference and (b) absorbance optics are 9.8 S and 9.9 S, respectively. The fits are based on data for dIgA1 at 0.96 mg/ml and a rotor speed of 20,000 r.p.m. The fits were determined from totals of 4-20 scans mid-way through the sedimentation experiment shown in Figure 5.7a,b.

5.2.4 Constrained modelling of the dIgA1 solution structure

Constrained modelling of the X-ray data was used to determine the solution structure of dIgA1. Two cycles of constrained molecular modelling fits were performed (Table 5.2). The dIgA1 modelling was initiated using the monomeric IgA1 solution structure model and the sequence and composition of dIgA1 (PDB code 1IGA; Boehm *et al.*, 1999). The monomeric IgA1 solution structure was assumed to be unchanged when incorporated into dIgA1. The 18 kDa J chain domain was initially not included in the dIgA1 constrained modelling because it comprises only 5.1% by mass of the dIgA1 structure, and it is located centrally in the dIgA1 structure where it will minimally perturb the scattering curve (Figure 5.1). These assumptions were each subsequently explored in control calculations (see below).

In the first cycle of fits, five modelling searches were performed. The generation of a sufficient number of trial dIgA1 models assumed that the two monomer IgA1 structures would be connected end-to-end through their Fc regions as shown schematically in Figure 5.1. This is consistent with the 26 nm length from the *P(r)* analyses (Figure 5.5a) and the dimensions of the IgA1 monomer (Section 5.2.2) and the known covalent linkages between the J chain and the tailpiece in each monomer. A dummy atom position was defined between the two monomers in order to act as the origin for the rotations and translations required to generate the dIgA1 models (Figure 5.9a; Section 5.4.4). The five searches each created 6,859 models from an initially-planar starting model, shown in Figure 5.9a. All five searches held Monomer 1 fixed, and rotated Monomer 2 about the dummy atom about the X-, Y- and Z-axes between 0° and 180° in 10° steps. Search 1 systematically explored all possible orientations between Monomers 1 and 2. Searches 2-5 repeated this to explore the effect of increasing the Fc-Fc separation between the IgA1 monomers. Searches 2-5 allowed for the possibility that the J chain acts as a spacer between the two IgA1 monomers, separating the two ends of the Fc regions. Thus in Searches 1, 2 and 3, the starting position of Monomers 1 and 2 were separated by 1.40 nm, 2.78 nm and 4.60 nm respectively between the base of the two Fc regions arranged coaxially with respect to each other (defined in Section 5.4.4). In Searches 4 and 5, while a separation of 2.78 nm was used again, Monomer 2 was now translated 2.0 nm and 4.0 nm respectively along the X-axis to explore asymmetry in

Filter for best-fit models	Models	Hydrated spheres (N)	X-ray R_G (nm)	X-ray R_{XS-1} (nm)	X-ray R_{XS-2} (nm)	Neutron R_G (nm)	X-ray R Factor	Neutron R Factor	$s^{\circ}_{20,w}$ (S)
Monomer orientation fits (fit cycle one)									
None	6859	2215-3700	6.51-9.01	1.87-5.37	0.93-2.16	5.97-7.94	5.5-20.6	5.9-14.7	n.a.
N, R_G (○ Figure 5.9)	1410	3426-3700	8.22-9.01	3.86-5.37	1.21-1.94	7.28-7.94	5.5-11.0	5.9-9.4	n.a.
N, R_G, R_{XS-1}, R_{XS-2}	54	3565-3681	8.22-8.52	3.92-4.33	1.36-1.51	7.30-7.53	5.5-7.5	7.5-9.3	n.a.
N, R_G, R_{XS-1}, R_{XS-2}, R Factor	10	3587-3649	8.22-8.52	4.22-4.33	1.42-1.51	7.34-7.53	5.5-5.9	7.5-8.0	9.76-9.89
Best-fit (Figure 5.11i)	1	3612	8.32	4.31	1.51	7.41	5.5	7.6	9.85
Fab conformational fits (fit cycle two)									
N, R_G, R_{XS-1}, R_{XS-2}, R Factor	10	3884-3933	8.70-8.89	3.78-4.11	1.58-1.70	7.49-7.63	5.5-5.7	7.3-8.0	8.84-9.03
Best-fit (Figure 5.16c; Figure 5.19)	1	3932	8.61	4.11	1.66	7.49	5.5	7.7	8.94
Experimental values	n.a.	3606	8.65 ± 0.27	3.94 ± 0.18	1.43 ± 0.07	7.60 ± 0.05	n.a.	n.a.	9.70 (SEDFIT) 9.85 (DCDT+)
Poor-fit dIgA1 models									
● (Figure 5.11j)	1	3616	7.87	4.13	1.57	7.12	10.6	9.7	10.17
▲ (Figure 5.11k)	1	3630	8.94	5.29	1.65	7.91	10.6	7.7	9.35
■ (Figure 5.11l)	1	3559	7.76	3.54	1.61	7.00	10.8	11	10.18

Table 5.2 Summary of the X-ray and neutron modelling fits for the dIgA1 solution structure

n.a. – data not available

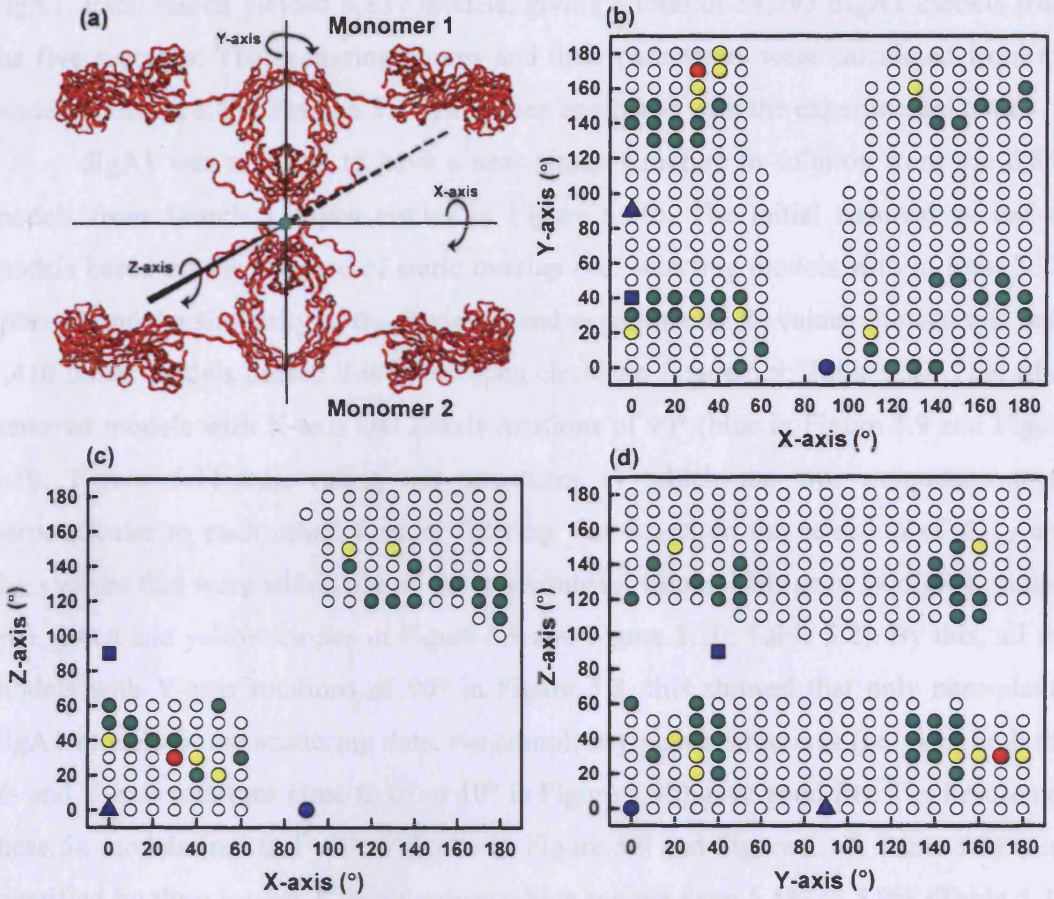

Figure 5.9 The starting model for the dIgA1 fit-cycle one searches and the orientations of the best-fit models for dIgA1

(a) The starting dIgA1 molecule (red) with a separation of 1.40 nm is shown in relation to the X-, Y- and Z-axes. The X- and Y-axes are in the plane of the paper, and the Z-axis is perpendicular to this. The dummy atom acting as the centre of rotation between Monomer 1 and Monomer 2 is shown in green. Monomer 1 is held fixed in the searches. (b, c, d) Projections in the XY, XZ and YZ planes are shown of the 1410 best-fit structures (○) from the 1.40 nm search that satisfy the absence of steric overlap and R_G filters (Table 5.2).The 54 (●) and 10 best-fit structures (●) are highlighted. The best-fit model corresponds to a X, Y, Z rotation of 30°, 170°, 30° (●). The poor fit models (●, ▲, ■) show X-, Y- or Z-axis rotations close to 90°.

dIgA1. Each search yielded 6,859 models, giving a total of 34,295 dIgA1 models from the five searches. The scattering curves and their parameters were calculated from the models (Section 5.4.4; Section 3.3.4) and then compared with the experimental data.

dIgA1 was revealed to have a near-planar structure in solution from the 6,859 models from Search 1 (open circles in Figure 5.10). The initial removal of bad-fit models based on the absence of steric overlap (i.e. selecting models with at least 3606 spheres) and the similarity of the modelled and experimental R_G values showed that only 1,410 dimer models passed this filter (open circles in Figure 5.9; Table 5.2). This filter removed models with X-axis and Z-axis rotations of 90° (blue in Figure 5.9 and Figure 5.10; Figure 5.11j,k,l), ruling out structures in which the two monomers were perpendicular to each other. Further filtering was based on the selection of $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ values that were within 5% of the experimental values. This gave 54 dIgA1 models (red, green and yellow circles in Figure 5.9 and Figure 5.10; Table 5.2). By this, all the models with Y-axis rotations of 90° in Figure 5.8, this showed that only near-planar dIgA1 models fit the scattering data. No completely planar structures (i.e. with both the X- and Y-axis rotations close to 0° or 10° in Figure 5.9b) gave good fits. The best ten of these 54 models (red and yellow circles in Figure 5.9 and Figure 5.10; Table 5.2) were identified by their lowest R factor values which ranged from 5.5% to 5.9% (Table 5.2). The best-fit dIgA1 model has the lowest R factor of 5.5% (red circles in Figure 5.9 and Figure 5.10; Table 5.2). This showed a near-planar structure with an X-, Y- and Z-axis rotation of 30°, 170° and 30° (Figure 5.11i; Figure 5.12). It gave visually good curve fits out to a Q value of at least 1 nm^{-1} for both the X-ray and neutron data, and reproduced the M and L values and the general appearance of the $P(r)$ curves (Figure 5.11a,e). The ten best-fit dIgA1 models all showed a small degree of non-planarity (X-axis rotations between 0° to 40°; Y-axis rotations between 150° to 20°; Z-axis rotations between 20° to 40°). All ten models show that the two Fc regions of the monomers are approximately coaxial, but exhibit a small degree of bend between the two monomers (Figure 5.12).

The best-fit dIgA1 structure accounted for the large difference between the X-ray and neutron R_G data. The superimposition of an α-carbon view of this model with both the unhydrated and hydrated sphere models showed that a larger proportion of the hydration spheres occupied voids within the Fab region (i.e. those between the two V

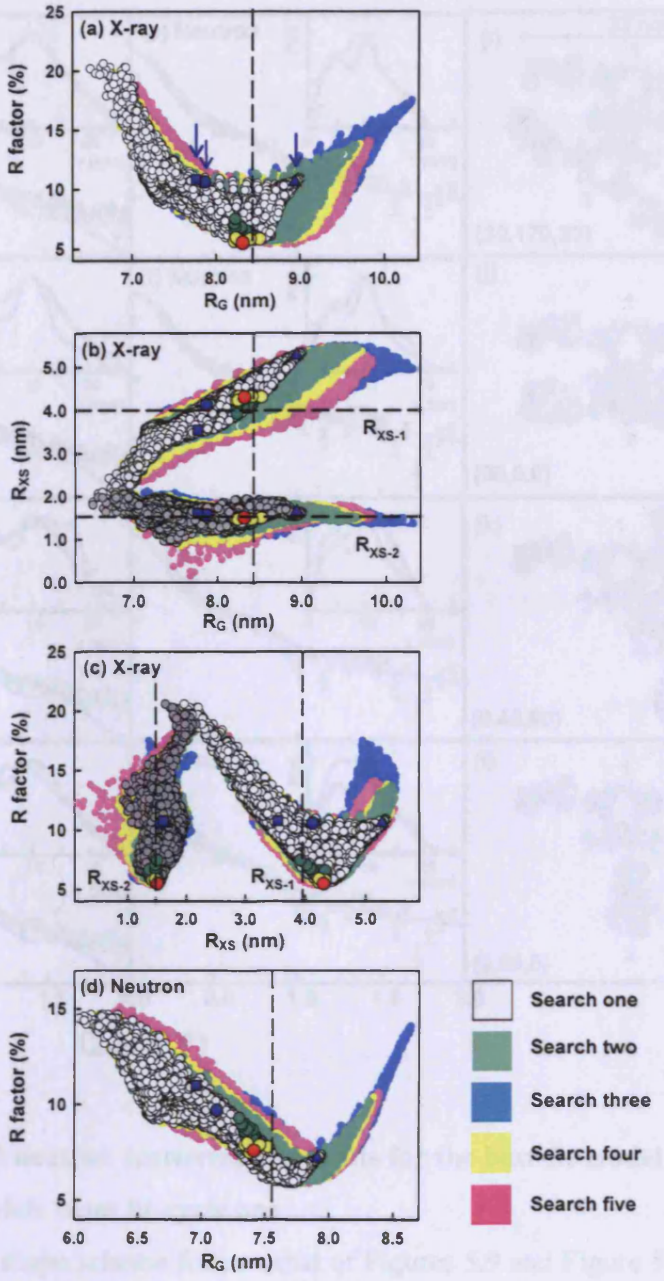

Figure 5.10 Outcome of the dIgA1 modelling from fit-cycle one

The parameters for 6859 rotational models of dIgA1 from the 1.40 nm search are shown as black open circles in all four panels. In Search 1, colour coding of the best-fit and poor-fit models follow that in Figure 5.9. The poor-fit models are arrowed in (a) for clarity. The results of Searches 2, 3, 4 and 5 are shown as green, yellow, pink and blue filled spheres respectively. The dashed lines indicate the experimental R_G, R_{XS-1} or R_{XS-2} values.

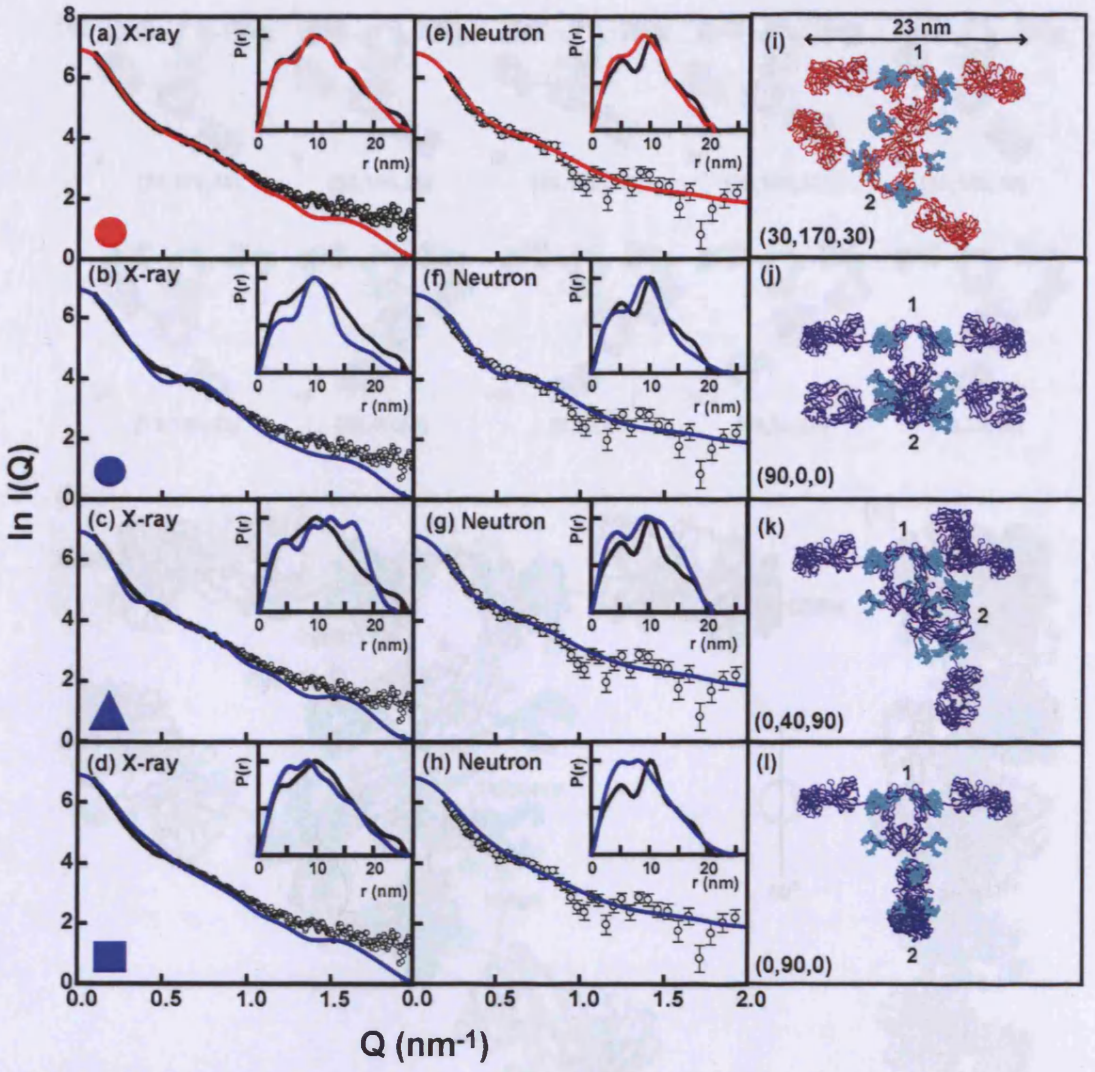

Figure 5.11 X-ray and neutron scattering curve fits for the best-fit model for dIgA1 and three poor-fit models from fit-cycle one

The colour coding and shape scheme follows that of Figures 5.9 and Figure 5.10. Panels (a,e,i) show the best-fit dIgA1 model from fit cycle one. The experimental *I(Q)* and *P(r)* curves are in black, and the red lines are the modelled best-fit curves. The *P(r)* curve is shown as an inset in the top right corner. The ribbon trace of this best-fit dIgA1 model is shown in (i), with the (X, Y, Z) rotation and the carbohydrate chains (cyan) shown. The tip-to-tip distance between the two Fab regions in an IgA1 monomer is arrowed in (i). Panels (b,f,j), (c,g,k) and (d,h,l) show the curve fits and structure of three poor fit models in which the two monomers are perpendicular to each other.

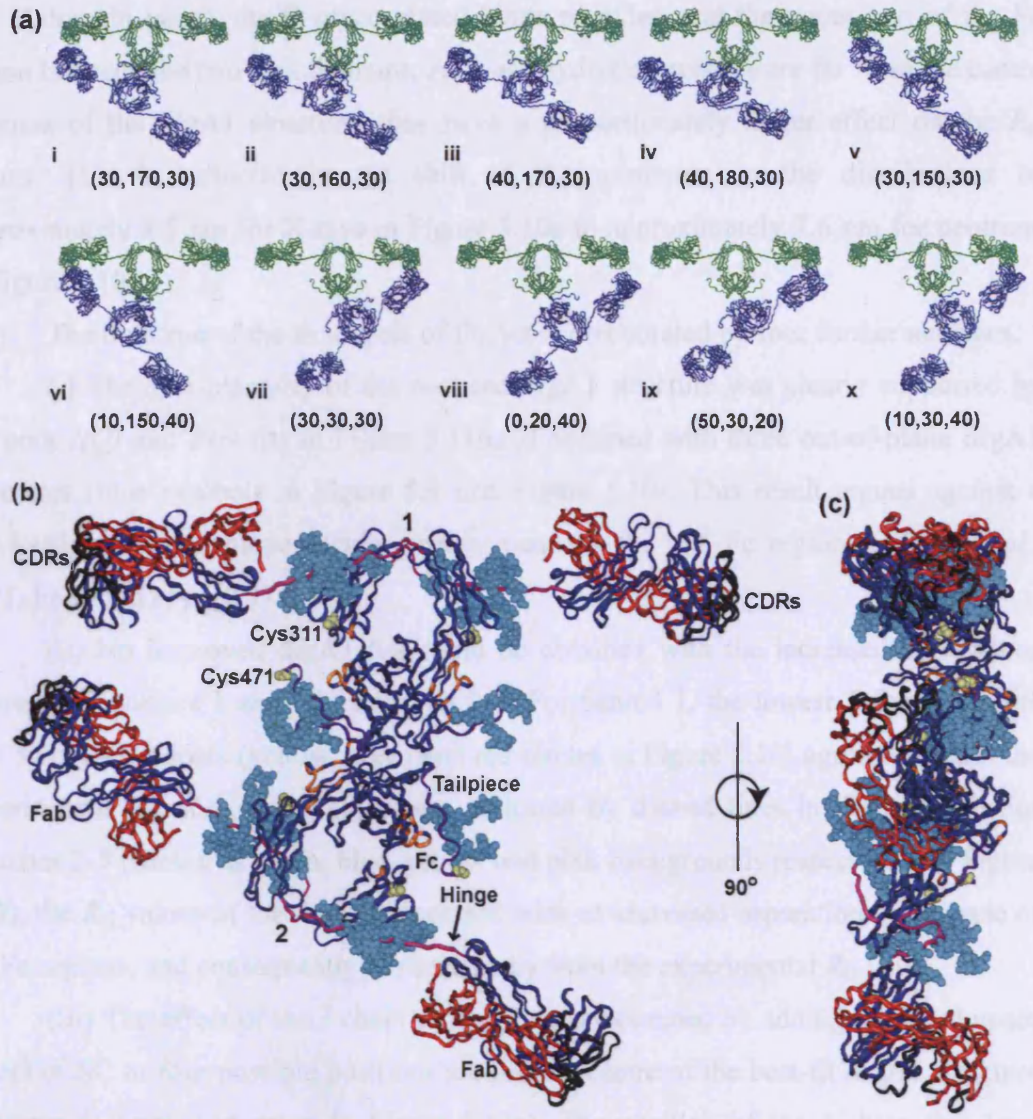

Figure 5.12 Best-fit models for dIgA1 from fit-cycle one

(a) The ten best-fit dIgA1 models from fit cycle one are shown projected on the XY plane, with the Fc-Fc regions in the same orientation in all ten views (● Figure 5.9; Table 5.2). Monomer 1 and 2 are shown in green and blue respectively. The best-fit dIgA1 model from Figure 5.11i is model (i). (b,c) Two orthogonal views of the best-fit model for dIgA1 from Figure 5.11i are shown. Heavy chains (blue), light chains (red), CDR residues (black), hinge regions (pink), tailpiece (purple) and residues in the C_H2 and C_H3 domains that bind to the FcαRI (orange) are shown. Cys311 in the C_H2 domain and Cys471 in the tailpiece are highlighted in yellow, and carbohydrate chains are shown in cyan.

and C domain pairs), the O-glycosylated hinge peptides, and the upper part of the Fc region between the two C_H2 domains. As these hydration spheres are far from the centre of mass of the dIgA1 structure, they have a proportionately larger effect on the R_G values. This is reflected in the shift of the minimum in the distributions of approximately 8.5 nm for X-rays in Figure 5.10a to approximately 7.6 nm for neutrons in Figure 5.10d.

The outcome of the first cycle of fits was corroborated by four further analyses.

(i) The near-planarity of the deduced dIgA1 structure was clearly supported by the poor *I(Q)* and *P(r)* fits in Figure 5.11b,c,d obtained with three out-of-plane dIgA1 structures (blue symbols in Figure 5.9 and Figure 5.10). This result argues against a previously-discussed perpendicular arrangement for the two Fc regions (Munn *et al.*, 1971; Feinstein *et al.*, 1971).

(ii) No improved dIgA1 fits could be obtained with the increased separations between Monomers 1 and 2 in Searches 2-5. For Search 1, the lowest *R*-factors of the best 54 dIgA1 mdoels (yellow, green and red circles in Figure 5.10) agree well with the experimental R_G, $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ values indicated by dashed lines in Figure 5.10. For Searches 2-5 (shown as green, blue, yellow and pink backgrounds respectively in Figure 5.10), the R_G values of the models increased with an increased separation at the base of the Fc regions, and consequently deviated away from the experimental R_G value.

(iii) The effect of the J chain on the fit was examined by adding the D4 domain model of SC to four possible positions around the centre of the best-fit dIgA1 structure (Chapter 4) (grey and green in Figure 5.13a). The position of the J chain that least affected the *R* factor of the best-fit model from Figure 5.11i was determined (green in Figure 5.13a). This J chain position was located proximate to two IgA1 tailpieces in the plane between two Fc regions, and agreed with the previous modelling of the IgM pentamer (Figure 1.7; Perkins *et al.*, 1991). As in the IgM modelling, it is stressed that both J chain model and its orientation within dIgA1 are arbitrary. There is no crystal structure for this, although an eight stranded beta-barrel structural model and a two-domain model have been proposed (Zikan *et al.*, 1985; Frutiger *et al.*, 1992; Section 1.3.3.1). The addition of the J chain to the best-fit dIgA1 model caused small reductions in the X-ray R_G, $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ values of 0.01-0.07 nm, and the *R* factor to increase by

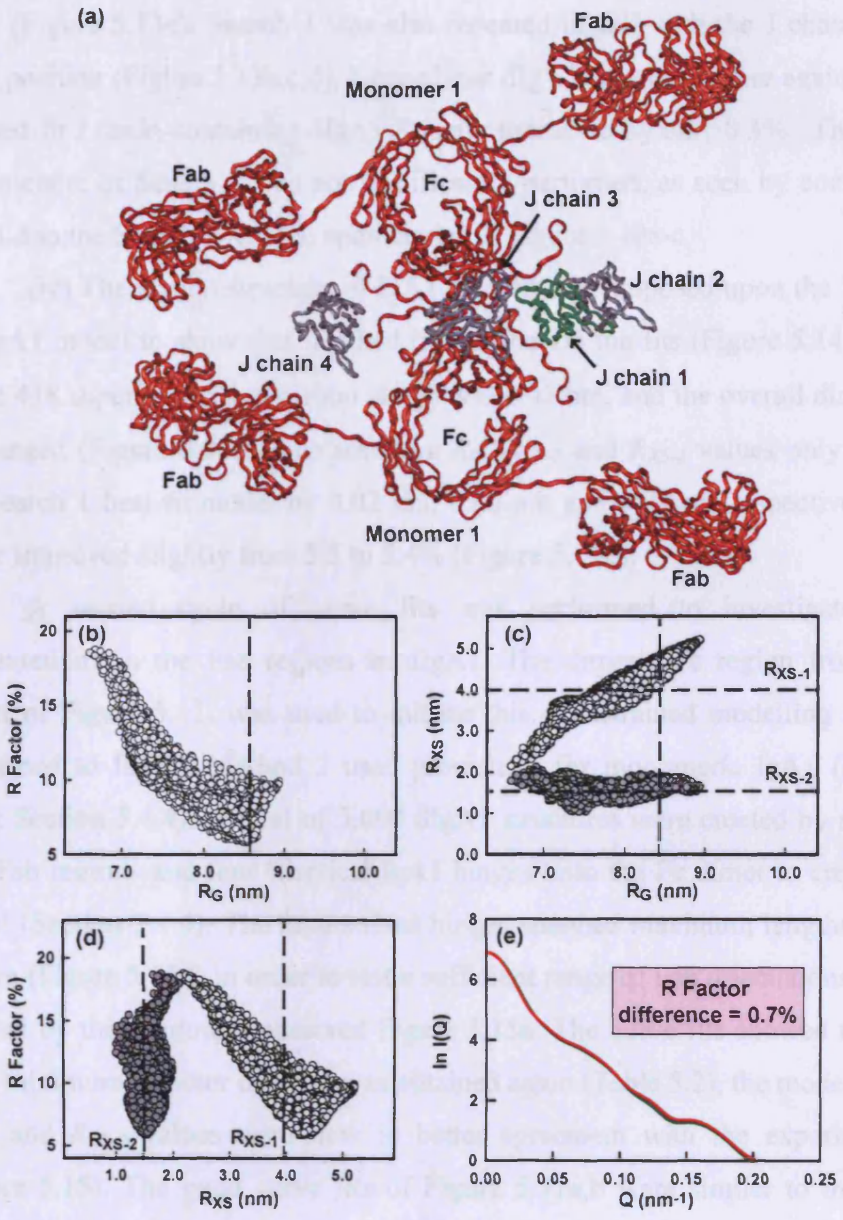

Figure 5.13 Effect of adding J chain to dIgA1 models from Search 1 of modelling fit-cycle one

(a) The best-fit dIgA1 structure from Figure 5.11i is shown in red with the four considered positions of the J chain also shown (J chain 1, green; J chains 2-4, grey). (b,c,d) The parameters for 6859 dIgA1 models with J chain in position 1 included are shown. Compare with Figure 5.10a,b,c. (e) The difference in the *R* factor (pink box) between the X-ray scattering curve of the best-fit dIgA1 model from Figure 5.11i with no J chain present (red line; Figure 5.11a) with the J chain present (green line) is 0.7%.

0.7% (Figure 5.13e). Search 1 was also repeated in full with the J chain added in the same position (Figure 5.13b,c,d). Near-planar dIgA1 structures were again obtained, and the best-fit J chain-containing dIgA1 *R* factor increased by only 0.3%. This showed that the outcome of Search 1 was not significantly perturbed, as seen by comparing Figure 5.12b-d to the Search 1 results, open circles in Figure 5.10a-c.

(iv) The crystal structure of IgA1 Fc was superimposed upon the Fc structure in the IgA1 model to show that this had little effect on the fits (Figure 5.14). The r.m.s.d. of the 418 superimposed α-carbon atoms was 0.43 nm, and the overall dimensions were unchanged (Figure 5.14a). The resulting R_G, $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ values only differed from the Search 1 best-fit model by 0.02 nm, 0.06 nm and 0.07 nm respectively, and the *R* factor improved slightly from 5.5 to 5.4% (Figure 5.14b).

A second cycle of curve fits was performed to investigate randomised reorientations in the Fab regions in dIgA1. The dimeric Fc region from the best-fit model of Figure 5.11i was used to initiate this. Constrained modelling searches were performed to follow Method 2 used previously for monomeric IgA1 (Boehm *et al.*, 1999; Section 5.4.4). A total of 3,000 dIgA1 structures were created by superimposing four Fab regions and four identical IgA1 hinges onto the Fc dimer to create the dIgA1 model (Section 5.4.4). The randomised hinges spanned maximum lengths of 6.5 nm to 9.0 nm (Figure 5.15d) in order to test a sufficient range of Fab orientations, and this was verified by the minimum observed Figure 5.15a. The curve fits showed that, while the same minimum *R* factor of 5.5% was obtained again (Table 5.2), the modelled X-ray R_G, $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ values were now in better agreement with the experimental values (Figure 5.15). The good curve fits of Figure 5.11a,b were similar to those of Figure 5.16a,b. The superimposition of the 32 best-fit models (green, yellow and red in Figure 5.15) showed that a small conformational family of best-fit structures had resulted in which the Fab regions displayed an extended and mostly T-shaped arrangement relative to their adjacent Fc region (Figure 5.17). All four Fab regions were conformationally independent of each other. When compared with the solution structure for monomeric IgA1 (Boehm *et al.*, 1999; Figure 1.8a), the Fab regions were again positioned in approximately the same plane as the Fc region. However the comparison of Figure 5.11i and Figure 5.16c shows that the Fab regions were displaced slightly upwards above the

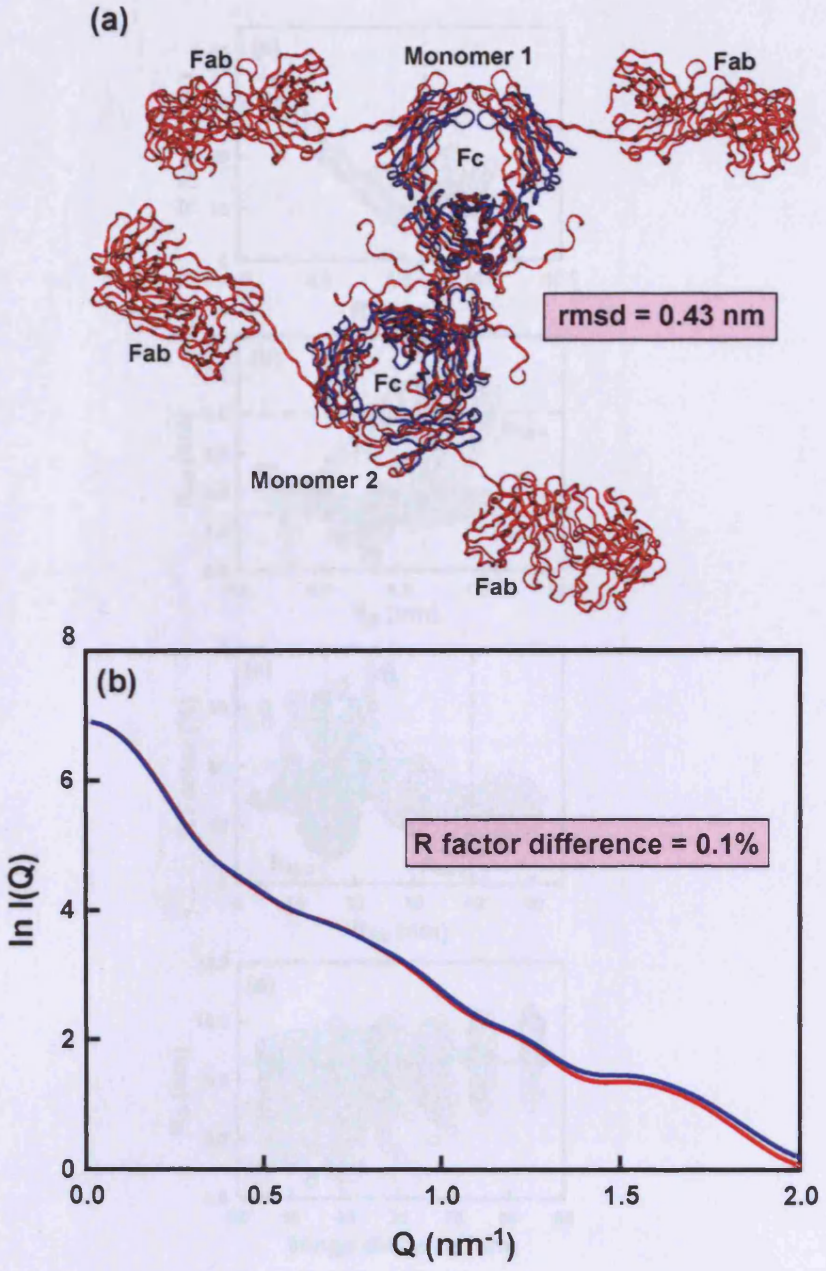

Figure 5.14 Effect of adding human IgA1 Fc crystal structure to the best-fit dIgA1 structure from fit-cycle one

(a,b) The best-fit dIgA1 model from Figure 5.11i (red) and the IgA1 Fc crystal structure (blue; PDB code 1OW0; Herr *et al.*, 2003b) are shown superimposed. (b) The X-ray scattering curve of the dIgA1 models is shown in the same colour scheme as (a). The superimposition r.m.s.d value and difference in *R* factors between the scattering curves are shown (pink boxes).

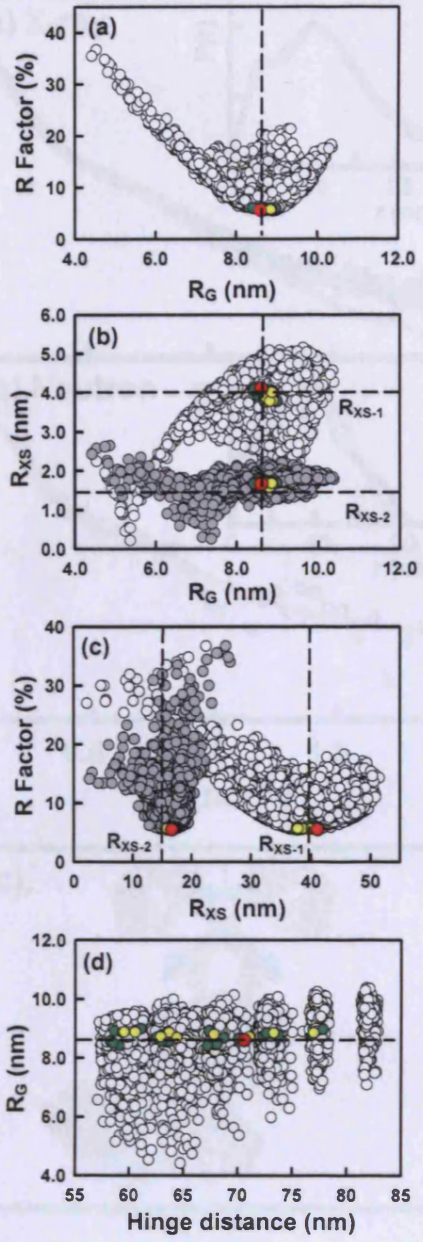

Figure 5.15 Outcome of the dIgA1 modelling from fit-cycle two

The parameters for 3,000 models of dIgA1 in which only the Fab regions and hinges are conformationally varied are shown as black open circles. The 32 (●) and 10 (●) best-fit structures are shown. The best-fit dIgA1 model (●) is shown (Table 5.2). The dashed lines indicates the experimental X-ray (a,b,d) R_G or (b,c) R_{XS-1} or R_{XS-2} values. (d) The dependence of the X-ray R_G values of the dIgA1 models on the hinge distance between the α-carbon atoms of Pro221 and His242 (Figure 5.1; Figure 5.2).

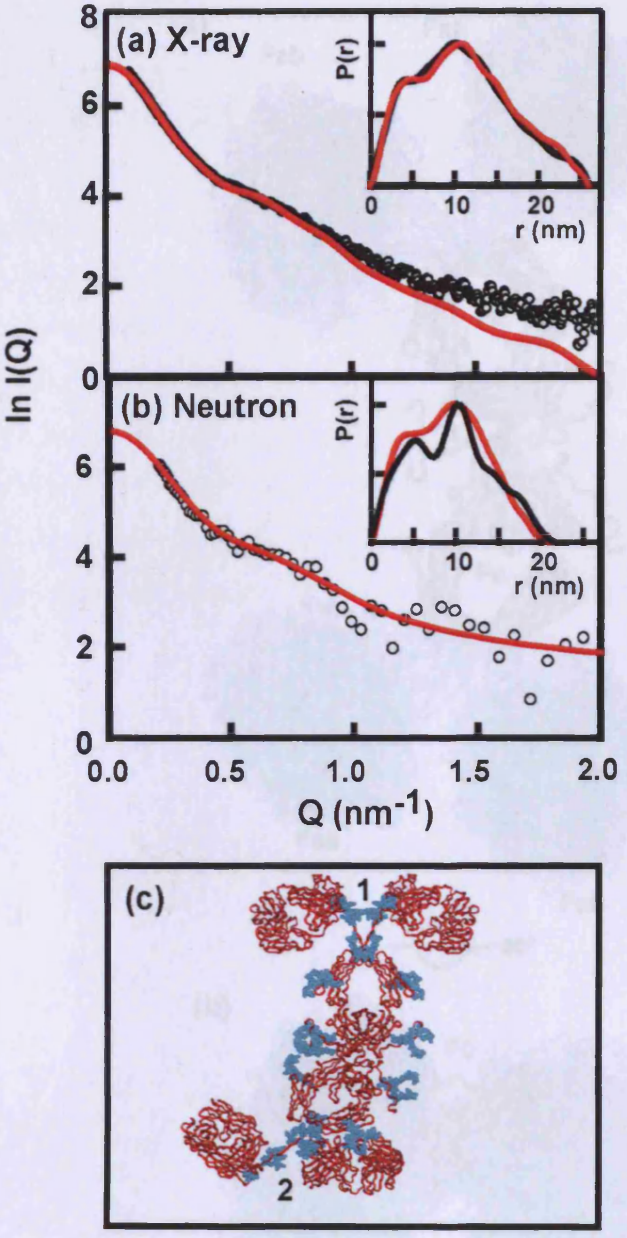

Figure 5.16 Scattering curves for the best-fit dIgA1 model from fit-cycle two

The (a) X-ray and (b) neutron experimental *I(Q)* and *P(r)* curves are in black, and the red lines are the modelled best-fit curves. The *P(r)* curve is shown as an inset in the top right corner. (c) The ribbon trace of this best-fit model (red) and the carbohydrate chains (cyan) are shown.

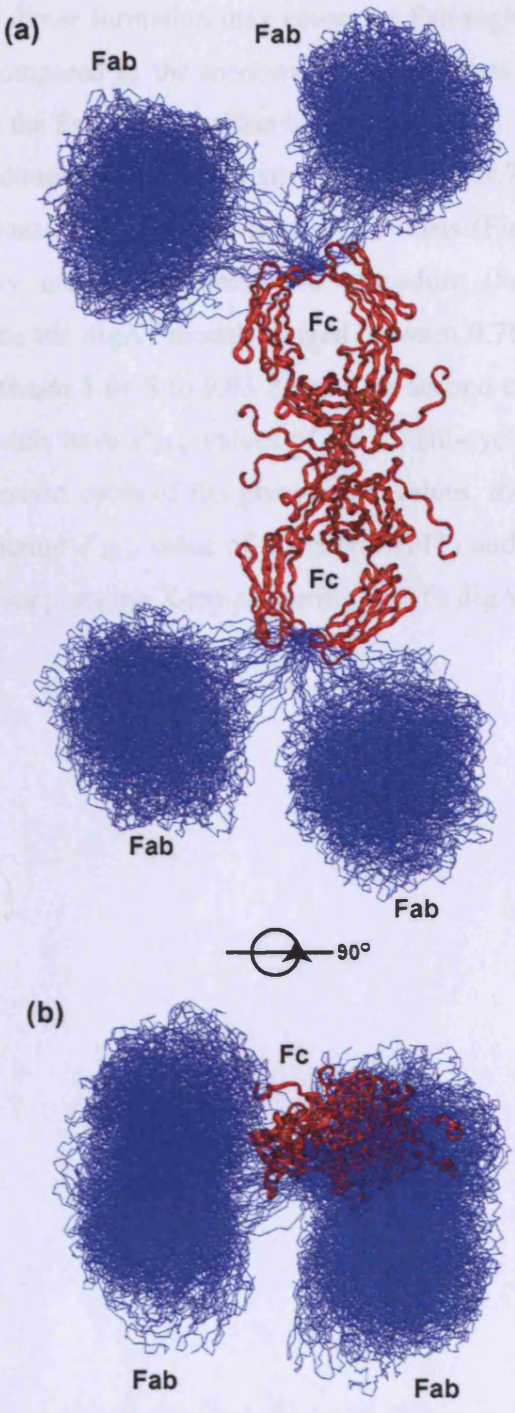

Figure 5.17 Superimposition of the 32 best-fit dIgA1 structures from fit-cycle two
(a,b) Two orthogonal views of the superimposition of the 32 best-fit dIgA1 models (● Figure 5.15) are shown. The Fc region is shown as a red ribbon at the centre of dIgA1, while the 32 sets of four Fab regions are shown as blue α-carbon traces.

Fc regions in dIgA1. Hence dimer formation may cause the Fab regions to move away from the Fc region when compared to the monomer. This suggests that the extended hinge conformation between the Fab and Fc region is semi-flexible.

The experimental sedimentation coefficient $s^0_{20,w}$ value of 9.7 S for dIgA1 was compared with the $s^o_{20,w}$ values of the ten best-fit dIgA1 models (Figure 5.12a; Figure 5.18). Using the previously established simulation procedure (Section 5.4.4), the calculated $s^0_{20,w}$ values for the ten dIgA1 models ranged between 9.76 S to 9.89 S from the first cycle of fits and between 8.84 S to 9.03 S from the second cycle of fits (Table 5.2). The best-fit dIgA1 models have $s^o_{20,w}$ values of 9.85 S (fit-cycle one) and 8.94 S (fit-cycle two). While the second cycle of fits gives lower values, the modelled values agree well with the experimental $s^o_{20,w}$ value of 9.7 S (SEDFIT) and 9.85 S (DCDT+) (Table 5.2). This agreement supports the X-ray scattering best-fit dIgA1 model of Figure 5.11i and Figure 5.16c.

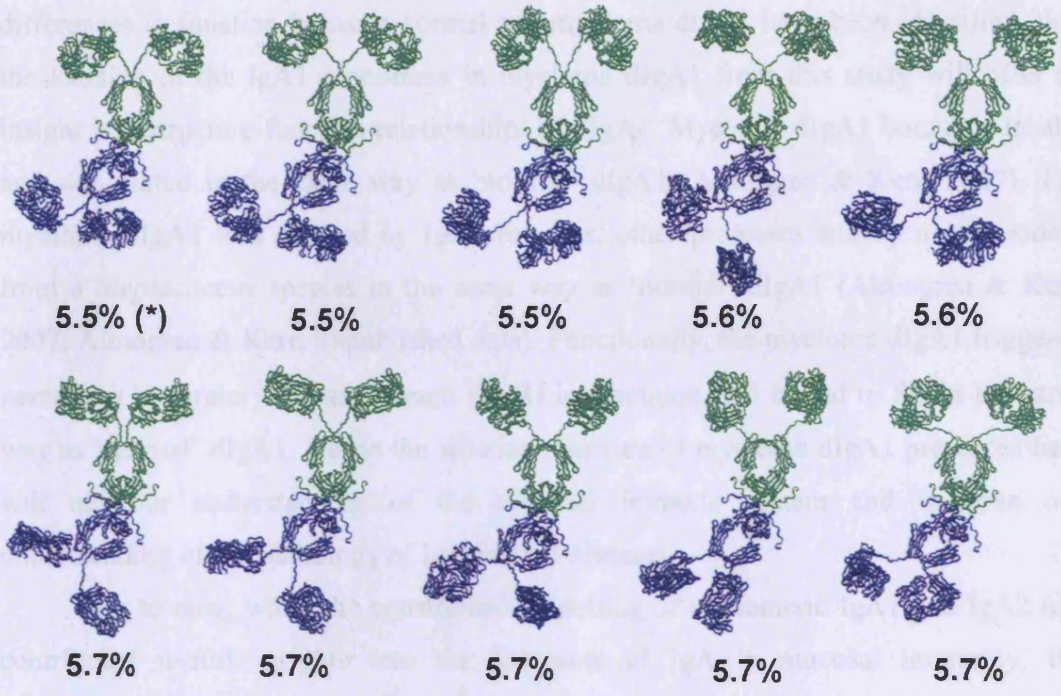

Figure 5.18 Ten best-fit dIgA1 structures from fit-cycle two

(a) The ten best-fit dIgA1 models from fit-cycle two (● in Figure 5.15) are shown projected onto the XY plane, with the Fc-Fc regions in each one shown in the same orientation in all ten views. Monomer 1 and 2 are shown in green and blue respectively. The first model (*) is the best-fit dIgA1 model from Figure 5.16c.

5.3 Discussion

dIgA is a building block of SIgA, one of the most important components of the immune system. dIgA in serum is less well characterised, although it is believed that an altered structural form of dIgA1 is central to the pathology of a number of IgA mediated diseases such as IgAN, the most common form of chronic glomerulonephritis. No differences in function between normal and myeloma dIgA1 have been identified, thus the location of the IgA1 monomers in myeloma dIgA1 from this study will offer an insight into structure-function relationships of dIgA1. Myeloma dIgA1 bound to Jacalin and was eluted in the same way as 'normal' dIgA1 (Almogren & Kerr, 2007). The myeloma dIgA1 was cleaved by IgA proteases, other proteases and by a glycosidase from a *Steptococcus* species in the same way as 'normal' dIgA1 (Almogren & Kerr, 2007; Almogren & Kerr, unpublished data). Functionally, the myeloma dIgA1 triggered neutrophil respiratory bursts through FcαRI interactions, and bound to SC in the same way as 'normal' dIgA1. Hence the solution structure of myeloma dIgA1 presented here will aid our understanding of the mucosal immune system and improve our understanding of the pathology of IgA-related diseases.

Up to now, while the constrained modelling of monomeric IgA1 and IgA2 had contributed useful insights into the functions of IgA in mucosal immunity, the equivalent information on polymeric IgA had yet to be presented. The size of dIgA1, its high carbohydrate content and its flexibility mean that the crystallisation of intact dIgA1 may not be realistic. Experimentally, we have shown here that dIgA1 has a near-planar structure in solution with the two monomers orientated end-to-end through their Fc regions. The two Fc regions are in proximity to each other, and are slightly bent in their relative arrangement (Figure 5.19). Because the dIgA1 structure was determined by fitting structural models against experimental scattering and ultracentrifugation data, this is not a prediction method and the final best-fit models qualify for deposition in the PDB.

While scattering modelling is not able to determine unique molecular structures, it is able to rule-out poor-fit models (Figure 5.11dj,k,l; Table 5.2). Both search cycles successfully converged on one small family of related best-fit structures after testing either all rotational orientations of Monomer 2 about Monomer 1, or all orientations of

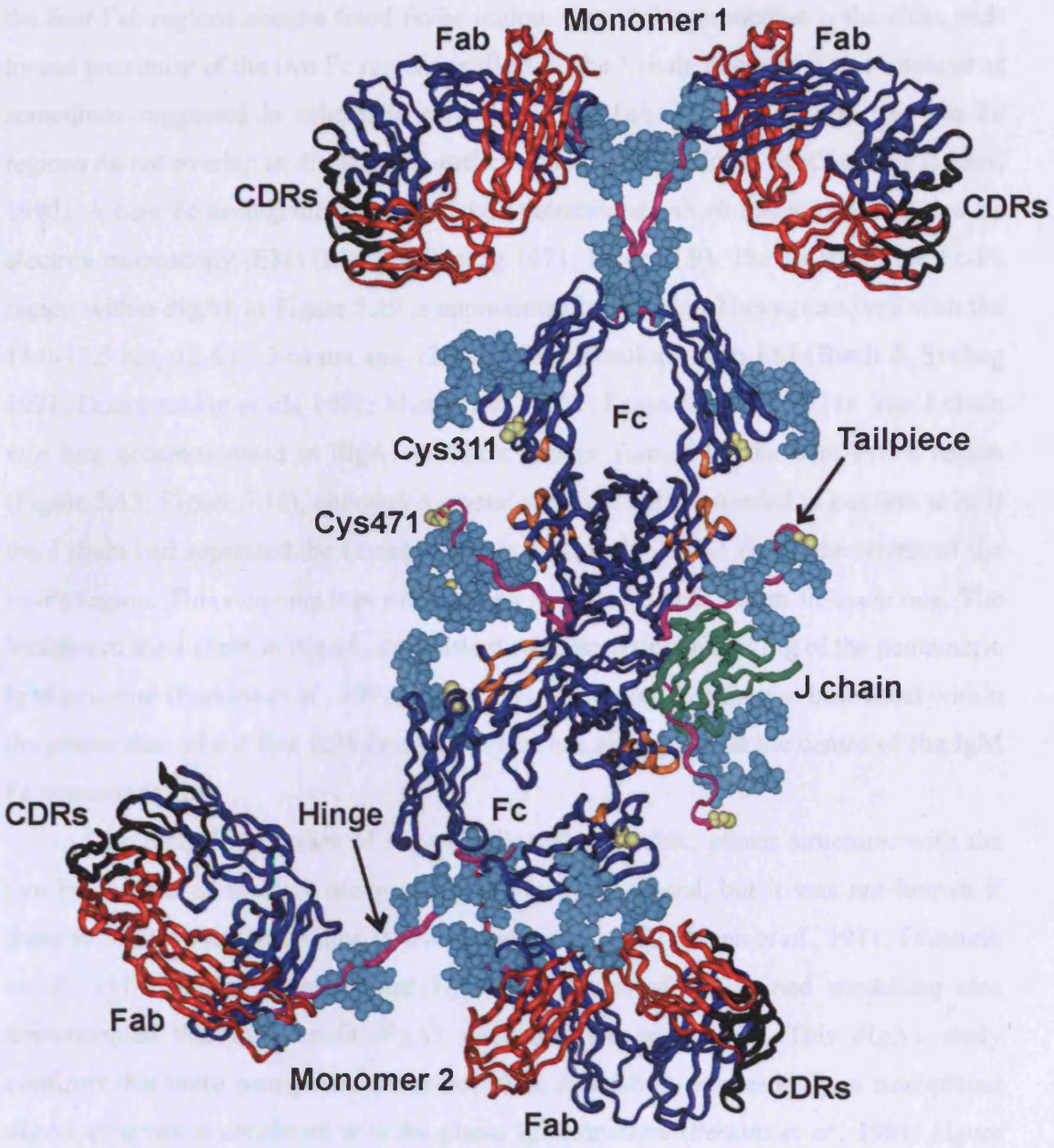

Figure 5.19 Molecular architecture of the dIgA1 solution structure

The best-fit dIgA1 structure from Figure 5.16 is shown. Heavy chains (blue), light chains (red) and the J chain (green) are shown. The hinge region and tailpiece (pink), the CDR residues (black) and the dIgA1 residues in the C_H2 and C_H3 domains that bind to the FcαRI (orange) are shown. Cys311 in the C_H2 domain and Cys471 in the tailpiece are highlighted in yellow, and carbohydrate chains are shown in cyan.

the four Fab regions about a fixed Fc-Fc region. One striking outcome is the close end-to-end proximity of the two Fc regions in dIgA1. The J chain does not act as a spacer as sometimes suggested in schematic cartoons of the IgA dimer. Likewise the two Fc regions do not overlap as discussed in earlier reviews (Mestecky & McGhee 1987; Kerr, 1990). A bent Fc arrangement in dIgA1 was determined, which has also been noted by electron microscopy (EM) (Bloth & Svehag 1971; Figure 1.9). The length of the Fc-Fc region within dIgA1 in Figure 5.19 is approximately 14.4 nm. This agrees well with the 14.0-15.5 nm, 13.8 (± 2.6) nm and 12.5 nm determinations from EM (Bloth & Svehag 1971; Dourmashkin *et al.*, 1971; Munn *et al.*, 1971; Feinstein *et al.*, 1971). The J chain was best accommodated in dIgA1 within a crevice formed by the bent Fc-Fc region (Figure 5.13; Figure 5.18), although a crystal structure will be needed to confirm this. If the J chain had separated the Fc regions, this would add 2.0-2.5 nm to the length of the Fc-Fc region. This outcome was ruled out by Searches 2 and 3 from fit-cycle one. The location of the J chain in dIgA1 is consistent with scattering modelling of the pentameric IgM structure (Perkins *et al.*, 1991; Figure 1.7). There, the J chain was best fitted within the planar disc of the five IgM Fc regions, and not as a spacer at the centre of the IgM Fc pentameric ring.

The early EM studies of dIgA1 indicated a flexible, planar structure, with the two Fc regions of the IgA monomers positioned end-to-end, but it was not known if these structures exist in solution (Dourmashkin *et al.*, 1971; Munn *et al.*, 1971; Feinstein *et al.*, 1971; Nezlin, 1998; Figure 1.7). Both cycles of constrained modelling also demonstrated that the best-fit dIgA1 structures are near-planar. This dIgA1 study confirms that these near-planar structures exist in solution. Interestingly, a near-planar dIgA1 structure is consistent with the planar IgM structure (Perkins *et al.*, 1991; Figure 1.7). dIgA1 models showing perpendicular arrangements of the two IgA1 monomers of the four Fab regions were eliminated by the constrained modelling filters. These perpendicular models resulted in visibly poorer *I(Q)* and *P(r)* curve fits (Figure 5.11b,c,d).

The solution structures for monomeric IgA1, the IgA1-HSA complex and dIgA1 structure all demonstrated similar T-shaped structures for the two Fab regions relative to the Fc region (Boehm *et al.*, 1999; Almogren *et al.*, 2006). A well-separated Fab and Fc

structure in IgA1 in all three cases was deduced from the observation of the *M2* peaks in the *P(r)* curves, meaning that the two Fab regions are seen, on average, to be extended away from the Fc region. Hinge flexibility of the Fab region relative to the Fc region is nonetheless expected, as observed from the variable Fab conformations seen in the EM studies of dIgA1 (Figure 1.9), the small shift in the Fab regions in the 32 best-fit models from fit-cycle two (Figure 5.17; Figure 5.18), and the slightly altered position of the Fab regions when going from monomer to dimer (Figure 1.8a; Figure 5.19). This is in accordance with the ability of IgA1 to bind efficiently to a range of Ags. The maintenance of these extended hinge conformations can be attributed to its primary structure (PVPSTPPTPSPPTPPTPSPSCCH) (Figure 5.2). Even though the hinge constitutes a comparatively long peptide of length 23 residues, the presence of the O-linked NeuNAc.Gal.GalNAc trisaccharides, multiple proline residues and the absence of glycine residues are expected to decrease the flexibility of the hinge and maintain their extended conformations.

Interestingly, human IgD also possesses an O-glycosylated segment within its hinge peptide of length 64 residues, and this was determined to have a similar T-shaped structure as that for IgA1 (Sun *et al.*, 2005). The solution structures of chimeric antibodies CR2-Ig and Crry-Ig also suggest why the hinge of dIgA1 may have reduced flexibility. Both chimeric antibodies contain a 39 residue hinge between the CR2 or Crry molecule and the Fc region of mouse IgG1. Due to the lack of hinge O-linked oligosaccharides, a low percentage of proline residues and no equivalent of immunoglobulin light chains being present these chimeric antibodies show increases in mobility about the hinge and do not exhibit two-fold symmetry (Gilbert *et al.*, 2006; Aslam *et al.*, 2003). As the four Fab regions are well separated within dIgA1, each Ag-binding site at the complementarity-determining regions (CDRs in Figure 5.19) is potentially able to interact with repeated epitopes independently of the three other Ag-binding sites with little steric interference. The valency of dIgA1 is therefore predicted to be four, and its avidity is predicted to be higher than that of the IgA1 monomer.

A near-planar dIgA1 structure would be well adapted for interactions with its receptors. By this, dIgA1 is able to approach the host cell membrane surface for effective interactions with a cell surface receptor, most notably FcαRI which is

responsible for IgA-mediated phagocytosis, oxidative burst and antibody-dependent cellular cytoxicity amongst other roles (Monteiro & van de Winkel, 2003). Receptor binding would occur independently of Ag binding. The dIgA1 residues involved in binding to FcαRI are highlighted in orange in Figure 5.19 (Herr *et al.*, 2003b; Pleass *et al.*, 1999). The superimposition of the Fc crystal structure complexed with two FcαRI molecules upon the dIgA1 model shows that the two FcαRI molecules project out of the plane of dIgA1 but on opposite sides of this dIgA1 plane (green in Figure 5.20). The superimposition of a second crystal structure onto the other monomer within dIgA1 shows likewise that the third and fourth FcαRI molecules likewise project outwards on both sides of this plane (yellow in Figure 5.20). Previously it has been shown that two FcαRI molecules can bind simultaneously to the monomeric Fc region in solution, and it has been speculated that this bivalent binding will occur at cell surfaces (Herr *et al.*, 2003a; Herr *et al.*, 2003b; Furtado *et al.*, 2004). Figure 5.20 likewise leads to the speculation that dIgA1 is able to bind up to two FcαRI molecules on a given cell surface, even though its predicted valency for the receptors is four. The spatial separation between the C-terminal α-carbon atoms of the two FcαRI receptors bound to monomeric IgA1 and dIgA1 is similar at 12.3 nm and 10.2 nm respectively, meaning that the same arrangement of FcαRI molecules will theoretically bind equally well to IgA1 monomers or dimers. However for reason of comparatively low binding affinities in relation to the IgA1 serum concentrations, it seems more likely that a 1:1 stoichiometry would result *in vivo* (Monteiro & van de Winkel, 2003; Woof *et al.*, 2005).

The appearance of dIgA1 as a near-planar structure with a bend between the two Fc regions leads to a possible mechanism for the formation of SIgA1 from dIgA1 and SC. While its position cannot be directly determined by scattering, the J chain is covalently bound to Cys471 in the tailpiece of the two monomers via its Cys69 and Cys15 residues, therefore joining the two monomers at the centre of the dimer. The bend between the two Fc monomers in the plane of IgA1 in the models of Figure 5.19 offers a covalently-bound location for the J chain within the crevice that is created by this bend. This location may enable the J chain in dIgA1 to be sufficiently exposed to bind to a central region in SC, this being the first event in the selective transportation of dIgA1

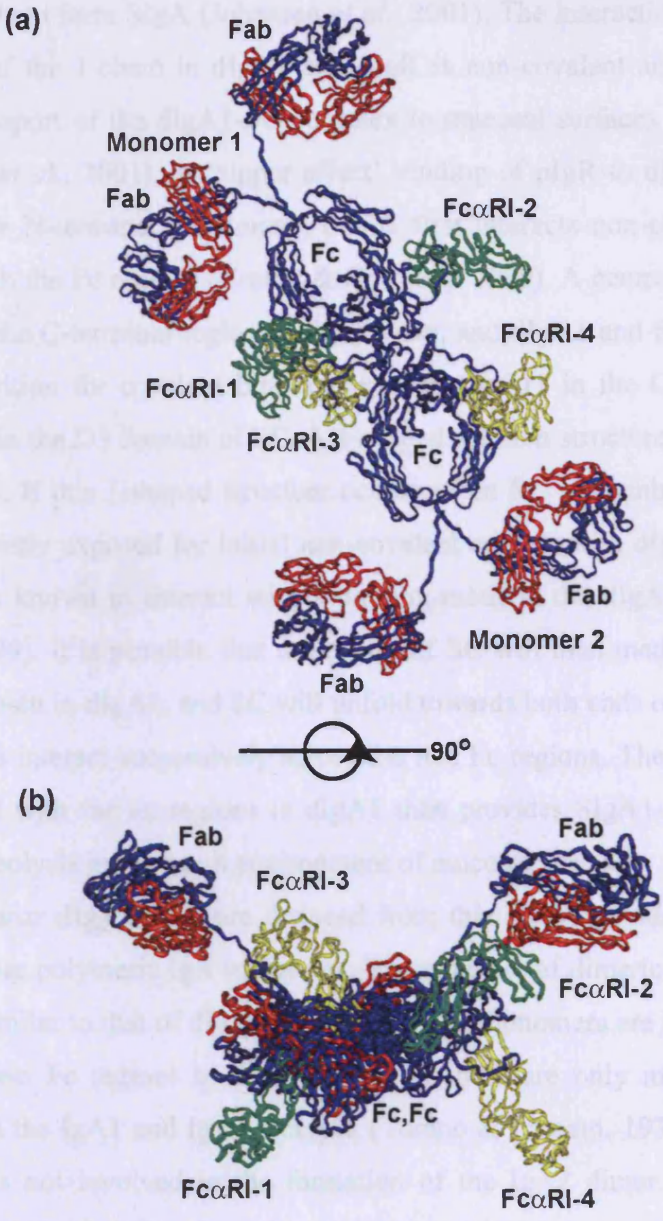

Figure 5.20 Interaction of dIgA1 with two FcαRI molecules

(a,b) The best-fit dIgA1 model of Figure 5.19 is shown after superimposition of the crystal structure of the human IgA Fc region complexed with two FcαRI molecules (PDB code: 1OW0; Herr *et al.*, 2003b) onto each Fc region in the dIgA1 model. The FcαRI molecules associated with Monomer 1 (green) and Monomer 2 (yellow) are labelled. The view in (b) corresponds to a rotation by 90° on the Y-axis of that in (a) to display the near-planarity of the dIgA1 structure and the projection of the FcαRI molecules above and below the dIgA1 structure, out of the dIgA1 plane.

across epithelial cells to form SIgA (Johansen *et al.*, 2001). The interaction between the C-terminal region of the J chain in dIgA1 and pIgR is non-covalent and represents a critical step for transport of the dIgA1-SC complex to mucosal surfaces and formation of SIgA (Johansen *et al.*, 2001). A 'zipper effect' binding of pIgR to dIgA1 has been proposed, where the N-terminal D1 domain of SC first interacts non-covalently with dIgA1 (possibly with the Fc region) (Crottet & Corthésy, 1999). A central region in SC then interacts with the C-terminal region of the J chain, and dIgA1 and SC are brought into the correct position for covalent binding between Cys311 in the C_H2 domain of dIgA1 and Cys502 in the D5 domain of SC. A J-shaped solution structure for free SC is shown in Chapter 4. If this J-shaped structure occurs when SC is membrane-bound as pIgR, D1 is sufficiently exposed for initial non-covalent contact with dIgA1 (Bakos *et al.*, 1991b). pIgR is known to interact with a binding-motif in the dIgA C_H3 domains (Hexham *et al.*, 1999). It is possible that the centre of SC will then mediate the initial contact with the J chain in dIgA1, and SC will unfold towards both ends of dIgA1 as the D1 and D5 domains interact successively across the two Fc regions. The interaction of the D1-D5 domains with the Fc regions in dIgA1 then provides SIgA1 with a greater ability to resist proteolysis in the harsh environment of mucosal surfaces.

The near-planar dIgA1 structure deduced from this study provides insight into the formation of other polymeric IgA molecules. The structure of dimeric IgA2 (dIgA2) is predicted to be similar to that of dIgA1 because the two monomers are joined through the base of their two Fc regions to the J chain and there are only minor sequence differences between the IgA1 and IgA2 isotypes (Torano & Putnam, 1978; Chapter 7). The hinge region is not involved in the formation of the IgA2 dimer. However the binding site of SC to dIgA2 differs from that to dIgA1, as SC binds covalently to dIgA1 but may not do so to dIgA2 (Almogren *et al.*, 2006). The two dIgA isotypes may therefore differ in a manner that remains to be elucidated, most probably mediated by the closer proximity of the Fab regions to SC in secretory IgA2. The antigenic reach of the two isotypes will differ for reason of the different hinge structures (Boehm *et al.*, 1999; Furtado *et al.*, 2004), and dimer formation is expected to accentuate this difference. The formation of trimeric and tetrameric IgA structures (Almogren & Kerr, 2007; Vaerman *et al.*, 1995) can be visualised by analogy with the solution structure

determination of planar pentameric IgM in which its Fc regions are disulphide linked through the Cys414 residues in the Cμ3 domains and the Cys575 residues in the tailpieces (Perkins *et al.*, 1991; Figure 1.7). Hexameric IgM structures without J chain have been described in which a sixth IgM monomer is inserted within this plane (Davis *et al.*, 1988). Cys311 in IgA1 (Figure 5.1; Figure 5.2) is located at the identical sequence and structural position to that of Cys414 in IgM (Herr *et al.*, 2003b). By analogy with IgM, It is relatively straightforward to insert further monomers within this Fc plane to form larger IgA oligomers. Using molecular graphics based on the IgA1 Fc crystal structure and the constrained scattering model for human IgM, it was possible to create trimeric (and tetrameric) assemblies of three (and four) co-planar Fc regions starting from the dimer with the J chain present, but only if the Fab regions are absent. If the Fab regions are present, they will readily lead to significant steric conflicts in the assembly of trimers and tetramers unless the Fab regions in question are significantly displaced out of the plane containing the three or four Fc regions.

IgAN is defined by the deposition of IgA in the glomerular mesangium of the kidney, and this is predominantly dIgA1 (Barratt & Feehally, 2005; Section 1.7.2). A high level of polymeric IgA1 such as in IgA myelomas does not necessarily lead to IgAN. A reduced galactosylation of the hinge O-linked oligosaccharides has been implicated in the pathology of IgAN. It is possible that this change may destabilise and alter the extended conformation of the IgA1 hinge peptide. Hence glycosylation changes may provide one possible mechanism for IgA aggregate formation by permitting the Fc regions to self-associate. Another possible mechanism for the formation of IgA deposition may arise from the observed consequences of the cleavage of the IgA1 hinge by bacterial proteases (Almogren & Kerr, 2007). This was seen to cause the cleaved Fc fragments derived from polymeric IgA1 (but not monomeric IgA1) to form disulphide-linked multimers (Almogren & Kerr, 2007). The molecular graphics modelling of planar Fc rings of IgA by analogy with IgM (see above) suggest that if the IgA1 hinges become more flexible through their deglycosylation and cause the Fab regions to move away from the Fc-Fc region, or the Fab regions are partially or completely removed by bacterial proteolysis, the near-planar assemblies of trimers, tetramers and higher oligomers of IgA can be generated starting from the double Fc region of dIgA1 and its J

chain. If this hypothesis is correct, these oligomers can be stabilised by disulphide bridges through Cys311 at the surface of the C_H2 domain in IgA, and may be responsible for the IgA deposits seen in IgAN.

The elucidation of a near-planar solution structure of dIgA1 has yielded a better understanding of the function of polymeric forms of IgA. This opens the way for the constrained modelling of SIgA1 and SIgA2. Knowledge of the monomer arrangement within the dimer has provided insights into the assembly of this dimer and higher oligomers of IgA1, and the way that this potentially interacts with the FcαRI receptor and pIgR and SC. However, many unsolved issues remain, including the exact binding motifs within dIgA1 that interact with SC in the formation of SIgA, and the structure and composition of pathological dIgA1 forms that are implicated in diseases such as IgAN.

5.4 Materials and Methods

5.4.1 Preparation and composition of dIgA1

A monoclonal dIgA1 with κ class light chain was isolated from a human myeloma serum using a combination of thiophilic chromatography and jacalin-agarose lectin affinity chromatography (Sun *et al.*, 2005). The human myeloma dIgA1 samples were prepared by Dr. Adel Almogren and Professor Michael A. Kerr. Samples were subjected to size-exclusion chromatography to remove non-specific aggregates and checked by reducing and non-reducing SDS-PAGE to confirm sample integrities before and after data collection. For X-ray scattering and AUC experiments, dIgA1 was dialysed into Dulbecco's PBS supplemented with EDTA and sodium azide as anti-bacterial preservatives (12.5 mM sodium phosphate, 140 mM NaCl, 0.5 mM EDTA, 0.02% NaN_3, pH 7.4). For neutron scattering, the buffer was Dulbecco's PBS as above, dialysed at 6°C into 100% ^2H_2O for 36 hours with four buffer changes. The amino acid sequence required for the dIgA1 data analyses was taken from the monomeric IgA1 solution structure (PDB code: 1IGA; Boehm *et al.*, 1999) which was derived from the SWISSPROT sequence code P01876 for the heavy chain constant domains and the crystal structure of human TR1.9 IgG1 Fab with a κ light chain (PDB code 1VGE; (Chacko & Padlan, 1996). The N-linked carbohydrates at Asn263 on the C_H2 domain and Asn459 on the tailpiece were each represented by a biantennary complex-type oligosaccharide with a $Man_3GlcNAc_2$ core and two NeuNAc.Gal.GlcNAc antennae. The five O-linked carbohydrates on each IgA1 hinge (Figure 5.1; Figure 5.2) were taken to be NeuNAc.Gal.GalNAc trisaccharides. The human J chain sequence was taken from the SWISSPROT sequence code P01591, to which another biantennary complex-type oligosaccharide was added at Asn49 (Figure 5.1; Figure 5.2). This composition of dIgA1 resulted in the following molecular parameters: molecular mass, 344,800 Da, an unhydrated volume of 439.4 nm^3, a hydrated volume of 580.2 nm^3, an absorption coefficient at 280 nm of 12.3 (1%, 1 cm), and a partial specific volume $\bar{v}$ of 0.724 ml/g.

5.4.2 X-ray and neutron scattering data collection and analysis

X-ray scattering data were obtained on the Beamline ID02 at the ESRF, Grenoble, France, collected by Dr. Patricia Furtado, with a ring energy of 6.0 GeV in

single-bunch mode with storage ring currents that ranged from 11 mA to 12 mA to reduce the incident X-ray flux (Narayanan *et al.*, 2001). The sample-to-detector distance of 3.0 m yielded a Q range from 0.07 nm^{-1} to 2.1 nm^{-1} (where $Q = 4 \pi \sin \theta / \lambda$; 2θ = scattering angle; λ = wavelength). Samples were contained in water-cooled Perspex cells at 15°C, of path thickness 1 mm and mica windows of thickness 25 μm at concentration of 0.58 mg/ml and 1.16 mg/ml. Samples were measured in 10 time frames, each of which lasted for 2 sec, these being optimal for the elimination of radiation damage effects.

Neutron scattering data were obtained on Instrument LOQ in two beam sessions by Dr. Patricia Furtado, at the pulsed neutron source ISIS at the Rutherford Appleton Laboratory, Didcot, U.K (Heenan *et al.*, 1997). The pulsed neutron beam was derived from proton beam currents of approximately 180 μA. Samples and buffers were measured in 2 mm-thick rectangular quartz Hellma cells positioned in a thermostatted rack at 15°C. Data acquisitions were lasted 4 to 13 h for dIgA1 samples at concentrations of 0.83 mg/ml and 1.16 mg/ml.

Other details, including calibrations and data reduction, the Guinier analyses to determine the radius of gyration R_G and the radius of gyration of the cross-sectional structure R_{XS}, and the calculation of the distance distribution function $P(r)$ using GNOM, are described previously in Section 2.2.4. For GNOM, the dIgA1 X-ray $I(Q)$ curve contained 524 data points between Q values of 0.98 to 2.02 nm^{-1} and was fitted with D_{max} set as 26 nm. The dIgA1 neutron $I(Q)$ curve contained 61 data points between Q values of 0.21 to 2.1 nm^{-1} and was fitted with D_{max} set as 23.5 nm.

5.4.3 Analytical ultracentrifugation data collection and analysis for dIgA1

SE data were acquired over 45 h using six-sector cells with column heights of 2 mm at rotor speeds of 5,000 r.p.m., 8,000 r.p.m., 11,000 r.p.m., 14,000 r.p.m., 17,000 r.p.m. and 20,000 r.p.m. at 20°C on a Beckman XL-I instrument equipped with an AnTi50 rotor. Absorbance scans at 280 nm and interference scans were recorded at six concentrations between 0.07 mg/ml and 0.97 mg/ml. Molecular mass values were determined on the assumption of a single species using ORIGIN v4.1 (Microcal). The buffer density was calculated to be 1.00543 g/ml and the viscosity was taken as 0.01002

cp. Sedimentation velocity data were acquired at concentrations of 0.53 mg/ml and 1.00 mg/ml over 16 h at rotor speeds of 10,000 r.p.m., 15,000 r.p.m. and 20,000 r.p.m in two-sector cells with column heights of 12 mm. Data were analysed using DCDT+ $g(s^*)$ time-derivative analyses and SEDFIT v9.4 based on the continuous *c(s)* distribution model with a resolution set as 150, while the cell meniscus and bottom, the frictional ratio of 1.638, and the partial specific volume $\bar{v}$, buffer density and viscosity were held fixed, and the baseline was allowed to float. Other details are as described previously in Section 2.2.4 and Section 4.4.3.

5.4.4 Constrained modelling of dIgA1

The full coordinates of monomeric IgA1 (PDB code 1IGA; Boehm *et al.*, 1999) was used to generate dimer models. The first cycle of fits created a planar starting model for the dimer, by arranging the two monomers Fc to Fc (Figure 5.9a). One IgA1 monomer was then rotated about a dummy atom created midway between the two Fc regions of the monomers, while the other monomer remained stationary. The J chain was not included. The rotations were in steps of 10° between 0° and 180° on each of the X-, Y- and Z-axes, hence this resulted in 19^3 or 6,859 models for each rotational search. Monomer separations were assessed by setting this as 1.40 nm, 2.78 nm and 4.60 nm in three searches in which the two starting monomers shared a common Y-axis. This separation was that between the two Cα atoms of Lys457 at the base of the two Fc regions, and the dummy atom is located at the mean position of all four Lys454 Cα atoms. In two final searches, the two IgA1 monomers had a separation of 2.78 nm, but the IgA1 monomer to be rotated was translated 2.0 nm and 4.0 nm along the X-axis.

The second cycle of fits randomised the hinge between the Fab and Fc regions in the best-fit dIgA1 structure obtained from the first modelling cycle. Six linker libraries with 5000 hinge conformations in each were generated using molecular dynamics (DISCOVER module of INSIGHT II, Accelrys) for 25-residue hinge lengths set between 5.5 nm to 9.0 nm in 0.5 nm steps, from each of which 500 conformations were randomly selected. The hinge between the C_H1 and C_H2 domains is 220-CPVPSTPPTPSPSTPPTPSPSCCHP-244 (Boehm *et al.*, 1999; Figure 5.2). Four identical hinges were used to generate the 3,000 dIgA1 models for evaluation.

In order to calculate the X-ray and neutron scattering curves for comparison with the experimental curves, each dIgA1 model was converted into Debye spheres. A cube side length of 0.531 nm with a cut-off of four atoms consistently gave models within 2% of their total unhydrated volume of 417.2 nm^3 (for that without the J chain as this had not been modelled), and gave a total of 2719 unhydrated spheres. After hydration, the number of hydrated spheres N was 3606. The values of N and the modelled R_G, R_{XS-1} and R_{XS-2} parameters were calculated in the same Q ranges used for the experimental Guinier analyses. Models that fit within 5% of N and the X-ray Guinier values were ranked using the goodness-of-fit R factor of the model. A flat-background correction of 0.5% of $I(0)$ was applied to the neutron fits to allow for residual incoherent scattering. The sedimentation coefficients $s^0_{20,w}$ were calculated from the coordinate models using HYDROPRO v7c. To take into account the hydration shell surrounding dIgA1, the recommended value of 0.31 nm for the atomic element radius for all atoms was used as an empirical approximation of this.

5.4.5 dIgA1 and pentameric IgM Protein Data Bank accession codes

The ten best-fit dIgA1 α-carbon co-ordinate models have been deposited in the Protein Data Bank with the accession code 2QTJ (Figure 5.18; Table 5.2). The α-carbon co-ordinate model of pentameric human IgM (Perkins *et al.*, 1991; Figure 1.7) was deposited with the accession code 2RCJ.

Chapter Six

Location of Secretory Component on the Fc Edge of Dimeric IgA1 Reveals Insight into the Role of Secretory IgA1 in Mucosal Immunity

6.1 Introduction

The mucosal surfaces, including the gastrointestinal, respiratory and urogenital tracts are the largest surface area in the human body in contact with the external environment (Brandtzaeg, 2007). Pathogenic microorganisms, their potentially toxic products and some components of ingested food and inhaled air are potential threats and are therefore antigenic at mucosal surfaces. Secretory immunoglobulin A (SIgA) is the principle Ab isotype present at mucosal surfaces where it mediates the first line of immune defence by neutralising and clearing the Ag (Nagler-Anderson, 2001). SIgA also restricts the indigenous commensal bacterial population entering the body from the lumen by immune exclusion (Corthésy, 2007). More IgA is produced daily than all the other Ab classes combined (Kerr 1990).

SIgA is produced locally in the mucosa, consisting of two (or sometimes three or four) 12-domain IgA monomers bound covalently through the joining (J) chain at the base of the Fc region (Figure 6.1). The transport of predominantly dimeric IgA (dIgA) across the epithelial lining into the luminal secretions, and the consequent formation of SIgA, is mediated by dIgA binding to the polymeric immunoglobulin receptor (pIgR) on the basolateral surface of epithelial cells. Following transcytosis of the complex across the cell and cleavage of pIgR at the luminal membrane, SIgA is released with the extracellular portion of the pIgR (now called secretory component (SC)) remaining bound to the IgA dimer (Figure 6.1). SC is not only necessary for the selective transportation and correct localisation and anchorage of SIgA, but it also stabilises markedly the structure of SIgA and increases its resistance to proteolysis in the harsh environment of mucosa (Crotett & Corthésy, 1998; Phalipon *et al.*, 2002; Lindh, 1975). SC binds covalently to dIgA by Cys502 to Cys311 in one of the C_H2 domains in an IgA1 monomer (Figure 6.1).

Despite the major role of SIgA1 in mucosal immunity, no new structural information on the assembly of this complex protein has been revealed since the electron microscopy studies from 38 years ago (Svehag & Bloth 1970; Bloth & Svehag 1971; Figure 1.9). The arrangement of the IgA1 monomers and SC within SIgA1 is as yet unknown. However solution structures for the IgA1 monomer and dIgA1 have been determined (Boehm *et al.*, 1999; Bonner *et al.*, 2008; Chapter 5). The IgA1 monomer

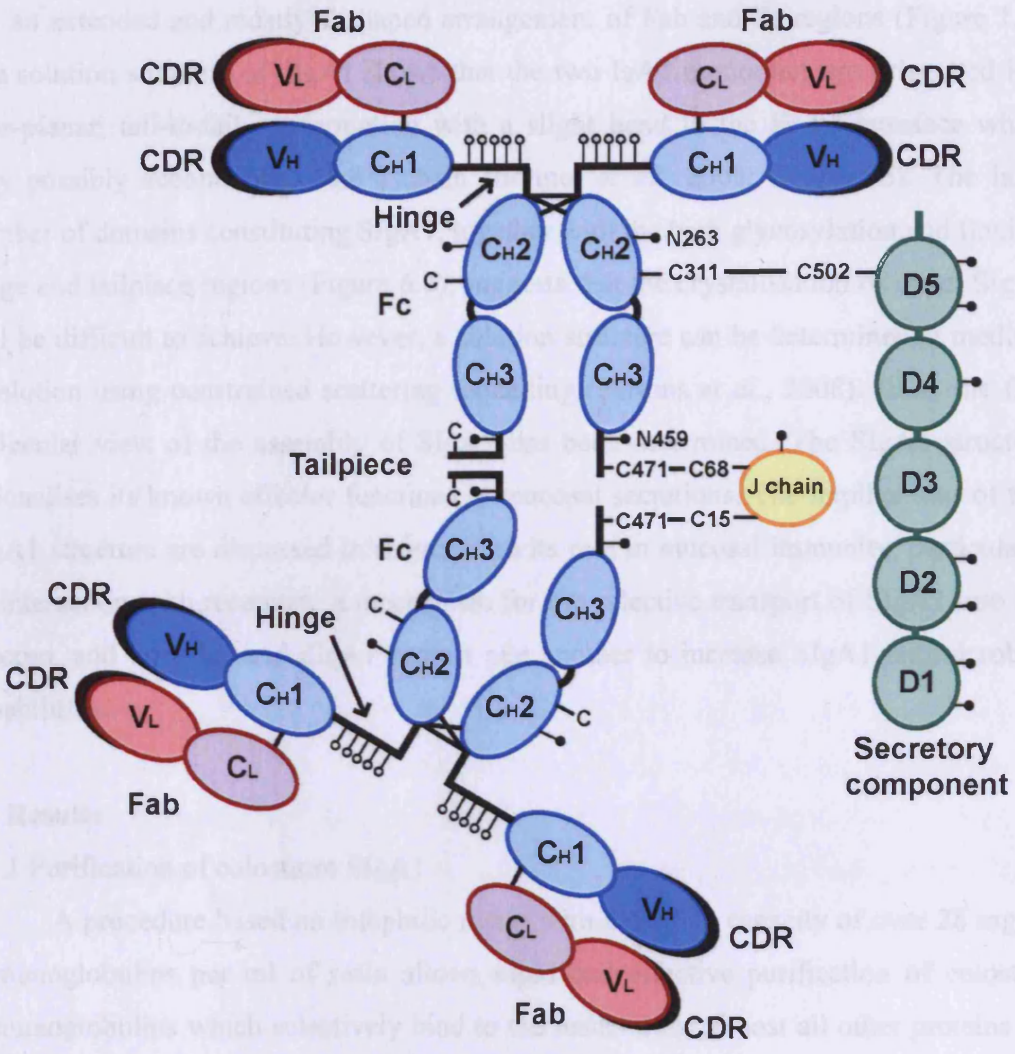

Figure 6.1 Domain structure of SIgA1

SIgA1 is composed of two IgA1 monomers covalently bound to the J chain (orange outline) and SC (green). Each IgA1 heavy chain (blue) contains the V_H, C_H1, C_H2 and C_H3 domains, and each light chain (red) contains the V_L and C_L domains. The complementarity-determining regions (CDR, black crescents), the 23-residue hinge and the 18-residue C-terminal tailpiece are labelled. The inter-heavy chain disulphide bridges at Cys241, Cys242, Cys299 and Cys301 are shown as an X. Cys471 in one tailpiece of each Fc region is covalently linked with Cys14 and Cys68 in the J-chain. Cys311 in one C_H2 domain is covalently linked with Cys502 in SC D5. The N-linked (●) and O-linked (○) oligosaccharide sites at Asn263 and Asn459 and the hinge regions are shown.

has an extended and mostly T-shaped arrangement of Fab and Fc regions (Figure 1.8). The solution structure of dIgA1 shows that the two IgA1 monomers are orientated in a near-planar, tail-to-tail conformation with a slight bend in the Fc-Fc interface which may possibly accommodate the J chain (Bonner *et al.*, 2008; Chapter 5). The large number of domains constituting SIgA1, together with the high glycosylation and flexible hinge and tailpiece regions (Figure 6.1), suggests that the crystallisation of intact SIgA1 will be difficult to achieve. However, a solution structure can be determined at medium resolution using constrained scattering modelling (Perkins *et al.*, 2008). Here, the first molecular view of the assembly of SIgA1 has been determined. The SIgA1 structure rationalises its known effector functions in mucosal secretions. The implications of this SIgA1 structure are discussed in reference to its role in mucosal immunity, particularly its interaction with receptors, a mechanism for the selective transport of SIgA1 into the mucosa, and how SC and dIgA1 protect one another to increase SIgA1 antimicrobial capabilities.

6.2 Results

6.2.1 Purification of colostrum SIgA1

A procedure based on thiophilic resins with a binding capacity of over 28 mg of immunoglobulins per ml of resin allows rapid and effective purification of colostral immunoglobulins which selectively bind to the resin whilst almost all other proteins do not (Sun *et al.*, 2005; Almogren *et al.*, 2007; Almogren & Kerr, 2007). Colostrum SIgA1 was then separated from SIgA2, free SC, secretory IgM and other colostrum proteins through its selective binding to a jacalin column. Samples were further purified by Sephacryl S-300 size-exclusion gel filtration, where dimeric SIgA1 eluted as a single symmetrical peak. After concentration, samples were checked by reducing and non-reducing SDS-PAGE and Western blot using an anti-IgA1 Fc-specific antibody. Reducing SDS-PAGE showed only the three expected bands for pure SIgA1 corresponding to the IgA1 heavy chain at 62 kDa, the light chain and J-chain which co-migrate at around 25 kDa and SC at around 80 kDa (Almogren *et al.*, 2007; Almogren & Kerr, 2007). The recovery rate for SIgA1 was in excess of 90%.

6.2.2 X-ray and neutron scattering of SIgA1

X-ray scattering data for SIgA1 were compared with previous scattering data for dIgA1 (Section 5.2.2) Comparison of *I(Q)* data for SIgA1 from a single time frame and those averaged over ten consecutive frames showed excellent signal-noise ratios with no radiation damage or SIgA1 aggregation, and only the first time-frames were used for analyses. At the lowest Q values ($Q = 4\ \pi\ \sin\ \theta/\lambda$; 2θ = scattering angle; λ = wavelength), the Guinier analyses resulted in linear plots, from which the R_G value (a monitor of macromolecular elongation) was obtained within satisfactory $Q.R_G$ limits (Figure 6.2a,b). The mean X-ray R_G value for SIgA1 was 8.29 ± 0.20 nm (6 values) (Figure 6.2a; Table 6.1). This was compared with previous scattering data for dIgA1 (Section 5.2.2) to indicate the extent to which the addition of SC (19% of the mass of SIgA1) to dIgA1 modifies the structure of the latter. The R_G for dIgA1 was 8.65 ± 0.27 nm (Table 6.1). The anisotropy ratio R_G/R_O (where R_O is the R_G value of the sphere with the same volume as the dry glycoprotein) for SIgA1 was 1.93. This is slightly less than that of 2.16 for dIgA1. These similarities between SIgA1 and dIgA1 suggest that SIgA1 has a similar elongated arrangement of its two IgA1 monomers in solution as that for dIgA1. The reduction in R_G/R_O suggested the SC is bound to the central region of dIgA1 if it is assumed that dIgA1 is unchanged in conformation.

The shorter axes of SIgA1 were monitored using X-ray cross-sectional Guinier analyses of *I(Q)* that resulted in the $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ parameters. The *ln(I(Q).Q)* plots revealed two linear regions in Q ranges between 0.20 – 0.28 nm^{-1} and 0.72 – 1.04 nm^{-1} that were similar to those for dIgA1 (Table 6.1). The mean $R_{XS\text{-}1}$ value of 3.90 ± 0.13 nm and mean $R_{XS\text{-}2}$ value of 1.27 ± 0.03 nm (5 determinations for both) (Figure 6.2c,d; Table 6.1) were close to those for dIgA1 of 3.94 ± 0.18 nm and 1.43 ± 0.07 nm respectively. This further indicated the similarity of the solution structures of SIgA1 and dIgA1.

Neutron scattering data acted as a control for X-ray-induced radiation damage, which does not occur with neutrons, any internal structural inhomogeneity caused by the 11% carbohydrate content of SIgA1, and any hydration effect which is largely invisible by neutrons but not by X-rays (Perkins, 2001). Data were obtained using a ^2H_2O buffer to correspond to a high negative glycoprotein-solvent contrast, in distinction to the high

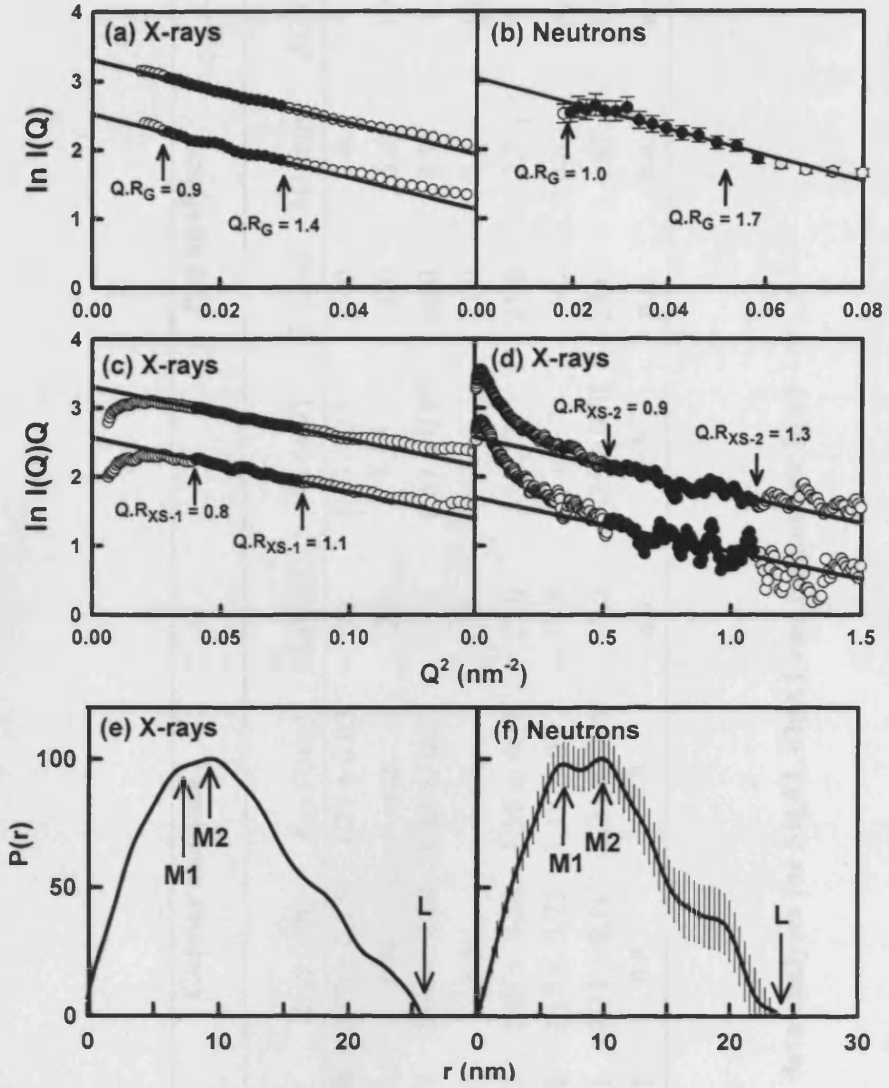

Figure 6.2 SIgA1 Guinier plots and distance distribution functions *P(r)* analyses

(a-d) The $Q.R_G$ and $Q.R_{XS}$ ranges used to determine the R_G, R_{XS-1} and R_{XS-2} values are represented by filled circles between the arrowed data points. The X-ray fits correspond to concentrations of 0.60 mg/ml (upper) and 0.30 mg/ml (lower). The neutron fit corresponds to data at 0.60 mg/ml. The X-ray and neutron Q ranges used for the R_G values were 0.11 to 0.17 nm^{-1} and 0.14 to 0.24 nm^{-1} respectively and that for the R_{XS-1} and R_{XS-2} values were 0.2 to 0.28 nm^{-1} and 0.72-1.04 nm^{-1} respectively. (e,f) The two most frequently occurring distances in the *P(r)* curve are denoted by the maxima *M*1 and *M*2 at 7.0 nm and 9.6 nm respectively (X-ray) and 6.8 nm and 10.0 nm respectively (neutron). The maximum length of SIgA1 is denoted by *L* at 26 nm (X-ray) and 24 nm (neutron) when *P(r)* reaches 0.

Protein	Data	Guinier analyses				*P(r)* analyses			
		R_G (nm)	$R_{XS\text{-}1}$ (nm)	$R_{XS\text{-}2}$ (nm)	L (nm)	R_G (nm)	L (nm)	$M1$ (nm)	$M2$ (nm)
Secretory IgA1	X-ray	8.29 ± 0.20	3.90 ± 0.13	1.27 ± 0.03	26	8.30 ± 0.17	26.0	7.0	9.6
	Neutrons	7.22	n.a.	n.a.	24	8.04	24.0	6.8	10.0
Dimeric IgA1	X-ray	8.65 ± 0.27	3.94 ± 0.18	1.43 ± 0.07	25.0	8.67 ± 0.17	26.0	4.9	9.9
(Chapter 5; Bonner *et al.*, 2008)	Neutrons	7.60 ± 0.05	n.a.	n.a.	22	7.47	23.0	5.1	10.1
Monomeric IgA1	X-ray	6.20 ± 0.13	2.20 ± 0.26	1.56 ± 0.16	21.0	6.12	21.0	3.7	9.1
(Boehm *et al.*, 1999)	Neutrons	6.11 ± 0.18	2.17 ± 0.23	1.18 ± 0.12	19.8	n.a.	n.a.	n.a.	n.a.
Monomeric IgA1-HSA	X-ray	7.47 ± 0.31	2.31 ± 0.11	1.31 ± 0.05	25.0	7.36 ± 0.01	25.0	4.87	8.25
(Almogren *et al.*, 2004)	Neutrons	7.09 ± 0.01	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

Table 6.1 Comparison of X-ray and neutron data analyses for SIgA1, dIgA1 and monomeric IgA1

All proteins are from human sources.

n.a. – not available

positive contrast observed by X-rays. The neutron R_G value for SIgA1 was 7.22 nm (1 determination) (Figure 6.2b; Table 6.1) which is similar to the dIgA1 neutron value of 7.6 nm. As for dIgA1, the neutron R_G value for SIgA1 is reduced by 1 nm compared to the X-ray R_G value, and this is attributable to a significant hydration shell positioned at large distances from the centre of the SIgA1 structure. The neutron anisotropy ratio R_G/R_O for SIgA1 was 1.85, which is similar to the X-ray-derived value above. This is again slightly less than the corresponding neutron R_G/R_O ratio of 2.08 for dIgA1. Neutron scattering also leads to the molecular mass from the linear relationship between the Guinier $I(0)/c$ values for proteins measured in ^2H_2O buffers when normalised relative to a standard deuterated polymer (where the molecular mass = $I(0)/c \times 9.10^5$) (Boehm *et al.*, 1999). The neutron Guinier $I(0)/c$ value of 0.43 resulted in a mass of 390 kDa for SIgA1 (Figure 2.5), which compares well with the sequence-derived value of 424 kDa from its composition which included four κ light chains, 16 N-linked biantennary and 20 O-linked oligosaccharides (Figure 6.1). The replacement of κ light chains by λ light chains in SIgA1 will result in a decrease of 7 kDa in mass.

The calculation of the distance distribution function $P(r)$ from the $I(Q)$ curve provides structural information in real space. The mean R_G values determined from the X-ray and neutron $P(r)$ curves were 8.30 ± 0.17 nm (6 values) and 8.04 nm (1 value) respectively, in good agreement with the Guinier analyses. The SIgA1 X-ray and neutron $P(r)$ curves reproducibly showed two maxima $M1$ and $M2$ located at 7.0 nm and 9.6 nm respectively (X-ray) and 6.8 nm and 10.0 nm respectively (neutron) (Figure 6.2e,f). These two peaks correspond to an abundance of inter-domain vectors found within the SIgA1 solution structure. While these were also observed for dIgA1 (Table 6.1), the X-ray $M1$ values at 6.8-7.0 nm for SIgA1 were larger than those of 4.9-5.1 nm for dIgA1, while the X-ray $M2$ values at 9.6 – 10.0 nm for SIgA1 were similar to those of 9.9-10.1 nm for dIgA1. The movement in $M1$ is attributed to the presence of bound SC in SIgA1. The maximum length, L, is determined from where the $P(r)$ curve reaches zero at larger r values. For SIgA1, L was determined to be 26 nm (X-ray) and 24 nm (neutron) (Figure 6.2e,f). These values agree well with those of 26 nm and 23 nm respectively for dIgA1, and provide further support to show that SIgA1 possesses a similar extended topology to that for dIgA1. These results show that the overall

structural differences between SIgA1 and dIgA1 are small.

6.2.3 SIgA1 analytical ultracentrifugation analyses

In order to confirm the molecular mass of SIgA1, sedimentation equilibrium (SE) experiments were performed at three concentrations and four rotor speeds (Section 6.4.2). Assuming that a single species was present, the individual SE fits for the molecular mass at each SIgA1 concentration and rotor speed gave values that ranged between 360 kDa to 420 kDa (Figure 6.3). Their extrapolation to zero concentration gave a mean molecular mass of 403 ± 43 kDa, which is within error of the composition-derived value of 424 kDa. This confirmed the neutron-derived molecular mass calculation (Section 6.2.3). The $s^0_{20,w}$ value independently measures macromolecular elongation, and acts as a control for the absence of polydispersity effects. Sedimentation velocity (SV) experiments were performed on SIgA1 (Section 6.4.2) and SEDFIT analyses gave a good visual agreement between experimental and fitted boundaries (Figures 6.4a,c). The resulting *c(s)* distribution plot revealed that a single SIgA1 species was present with an $s°_{20,w}$ value of 10.8 S from both the interference and absorbance data (Figures 6.4b,d). No other species were detectable and only dimeric SIgA1 was detected, as tetrameric SIgA has an $s°_{20,w}$ value of 15.5 S (Mestecky *et al.*, 2005). The conversion of the *c(s)* plots to *c(M)* mass distributions showed that the 10.8 S species corresponded to a molecular mass of 440 kDa (data not shown), which is consistent with the neutron, SE and sequence-derived molecular masses above. The *c(s)* analyses were confirmed by the *g(s*)* time derivative analyses using DCDT+ of between 6-14 scans at the centre of the sedimentation profile. These gave an $s°_{20,w}$ value of 10.8 S (interference) and 10.9 S (absorbance) (Figure 6.4e,f). The new determinations agree well with a previous $s°_{20,w}$ value of 10.9 S (Bjork & Lindh, 1974) and with six earlier values between 10.5 S to 11.7 S previously reviewed (Heremans, 1974). The frictional ratio f/f_o was determined from $s^0_{20,w}$ value, where f_o is the frictional coefficient of a sphere with a volume equal to that of the hydrated macromolecule. The frictional ratio of SIgA1 is 1.70, which is similar to 1.63 for dIgA1 (Section 5.2.3). This comparison showed that SIgA1 and dIgA1 show similar elongated structures in solution, in agreement with the previous ultracentrifugation analyses and the new scattering results.

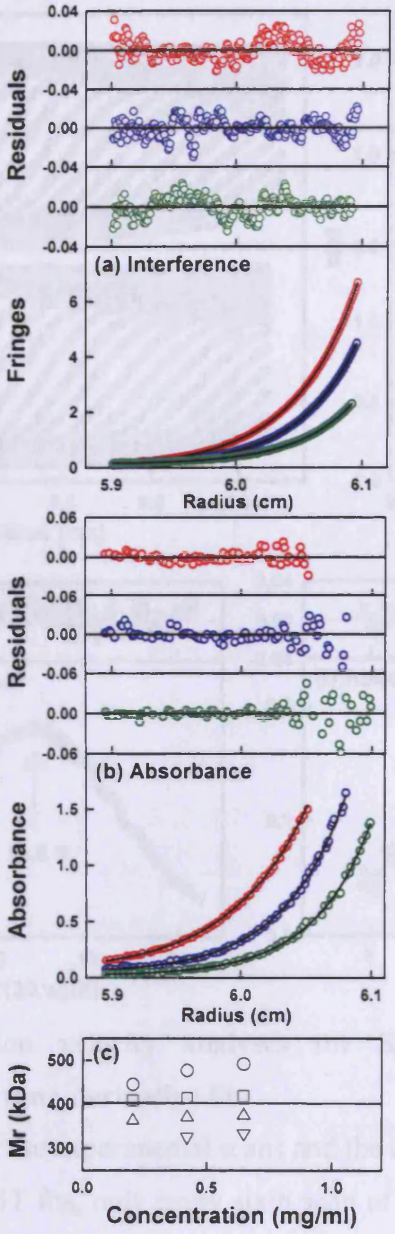

Figure 6.3 Sedimentation equilibrium analyses for SIgA1

SIgA1 was analysed using (a) interference and (b) absorbance optics at concentrations of 0.65 mg/ml (red), 0.43 mg/ml (blue), and 0.22 mg/ml (green). Circles represent the experimental data at a rotor speed of 8,000 r.p.m. and the continuous black lines represent their fits. The corresponding curve fit residuals are shown above the exponential fits. (c) The rotor speed dependence of the fitted molecular mass gave a mean value of 403 ± 10 kDa, denoted by (○) 5,000 r.p.m, (□) 8,000 r.p.m, (△) 11,000 r.p.m and (▽) 14,000 r.p.m.

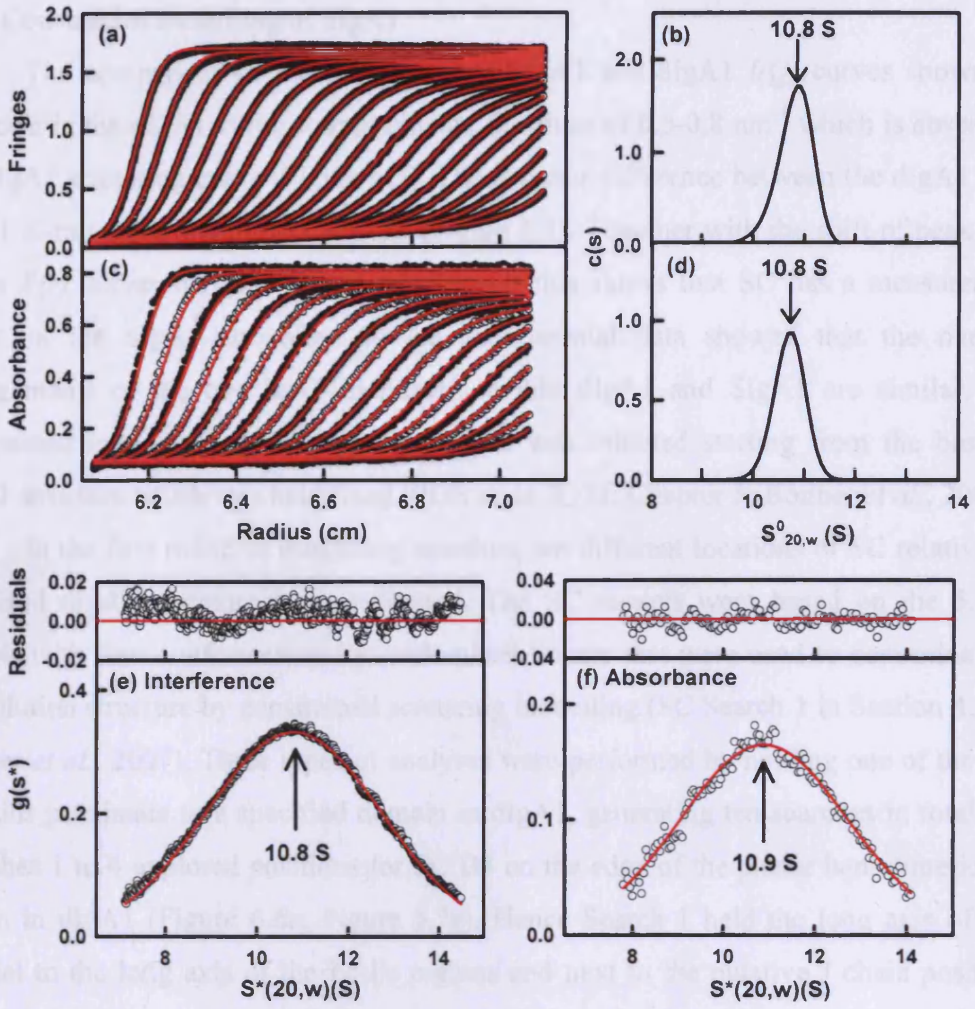

Figure 6.4 Sedimentation velocity analyses for SIgA1 using SEDFIT *c(s)* distribution and DCDT+ time-derivative fits

The black circles represent the experimental scans and the continuous red lines represent the fits. (a,c) In the SEDFIT fits, only every sixth scan of the 120 scans used is shown for reason of clarity. SIgA1 was measured using (a) interference and (c) absorbance optics at a concentration of 0.58 mg/ml and a rotor speed of 20,000 r.p.m. (b,d) The corresponding *c(s)* plots from (a,c) are shown from which the sedimentation coefficient for SIgA1 was determined to be 10.8 S. (e,f) DCDT+ *g(s*)* fits for SIgA1 at 0.58 mg/ml and at a rotor speed of 20,000 r.p.m. The goodness-of-fit residuals are shown above each analysis. The arrowed $s_{20,w}$ values using (e) interference and (f) absorbance optics are 10.8 S and 10.9 S, respectively. The fits were determined using between four to 14 scans recorded mid-way through the experiment shown in (a,c).

6.2.4 Constrained modelling of SIgA1

The comparison of the superimposed SIgA1 and dIgA1 *I(Q)* curves shows an inflexion in the dIgA1 curve at approximate *Q* values of 0.5-0.8 nm^{-1} which is absent in the SIgA1 scattering curve (Figure 6.5). The *R* factor difference between the dIgA1 and SIgA1 X-ray scattering curves is 8.5% (Figure 6.5). Together with the shift of peak *M1* in the *P(r)* curves between SIgA1 and dIgA1, this shows that SC has a measureable effect on the SIgA1 structure. As the experimental data showed that the overall arrangements of the two IgA1 monomers within dIgA1 and SIgA1 are similar, the constrained modelling of the SIgA1 structure was initiated starting from the best-fit dIgA1 structure which was held fixed (PDB code 2QTJ; Chapter 5; Bonner *et al.*, 2008).

In the first round of modelling searches, ten different locations of SC relative to the fixed dIgA1 structure were evaluated. The SC models were based on the 5,000 models with four conformationally-randomised linkers that were used to determine the SC solution structure by constrained scattering modelling (SC Search 1 in Section 4.2.5; Bonner *et al.*, 2007). Three types of analyses were performed by holding one of the SC domains proximate to a specified domain in dIgA1, generating ten searches in total. (i) Searches 1 to 4 explored positions for SC D3 on the edge of the planar bent dimeric Fc region in dIgA1 (Figure 6.6a; Figure 6.7a). Hence Search 1 held the long axis of D3 parallel to the long axis of the Fc-Fc regions and next to the putative J chain position within the Fc-Fc crevice. Search 2 repeated this with D3 now rotated by 90° to be perpendicular in position to that used in Search 1. Searches 3 and 4 repeated these searches on the other edge of the Fc-Fc regions. (ii) Searches 5 and 6 explored locations for SC on top of the face of the planar Fc-Fc regions (Figure 6.6a; Figure 6.7a). Searches 5 and 6 positioned the long axis of D3 either parallel or perpendicular to the long axis of the Fc-Fc regions. Tentative models of this type where SC is wrapped around the planar Fc-Fc region are predicted in reviews but have no experimental basis (Kerr, 1990; Mestecky *et al.*, 2005; Brandtzaeg, 2007; Heremans, 1974; Corthésy, 2003; Mestecky & McGhee, 1987; Woof & Kerr, 2006). (iii) Searches 7 to 10 explored SC models based on the knowledge that Cys502 in D5 forms a disulphide bridge with Cys311 in one of the four C_H2 domains (Figure 6.1; Figure 6.6a; Figure 6.7a; Fallgreen-Gebauer *et al.*, 1993). Accordingly D5 of SC was positioned next to each of the four

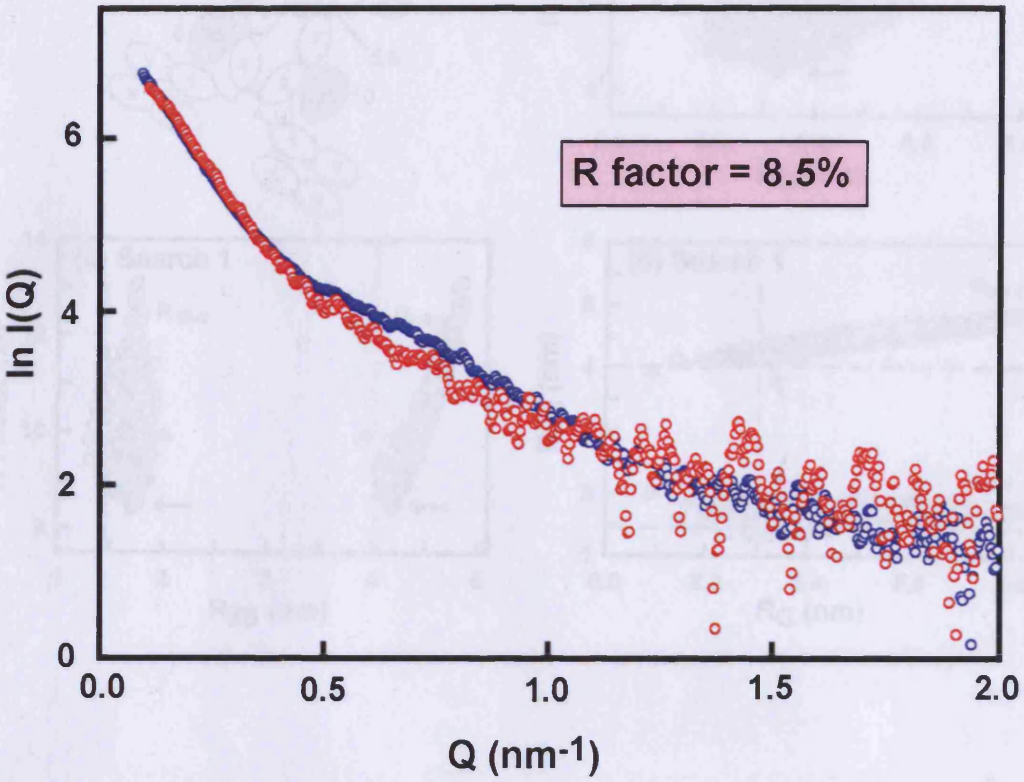

Figure 6.5 Comparison of the SIgA1 and dIgA1 experimental *I(Q)* scattering curves

The difference between the SIgA1 experimental X-ray scattering curve (red) and the dIgA1 experimental X-ray scattering curve (blue; Figure 5.10; Figure 5.15) is shown. The *R* factor of the two curves is 8.5%.

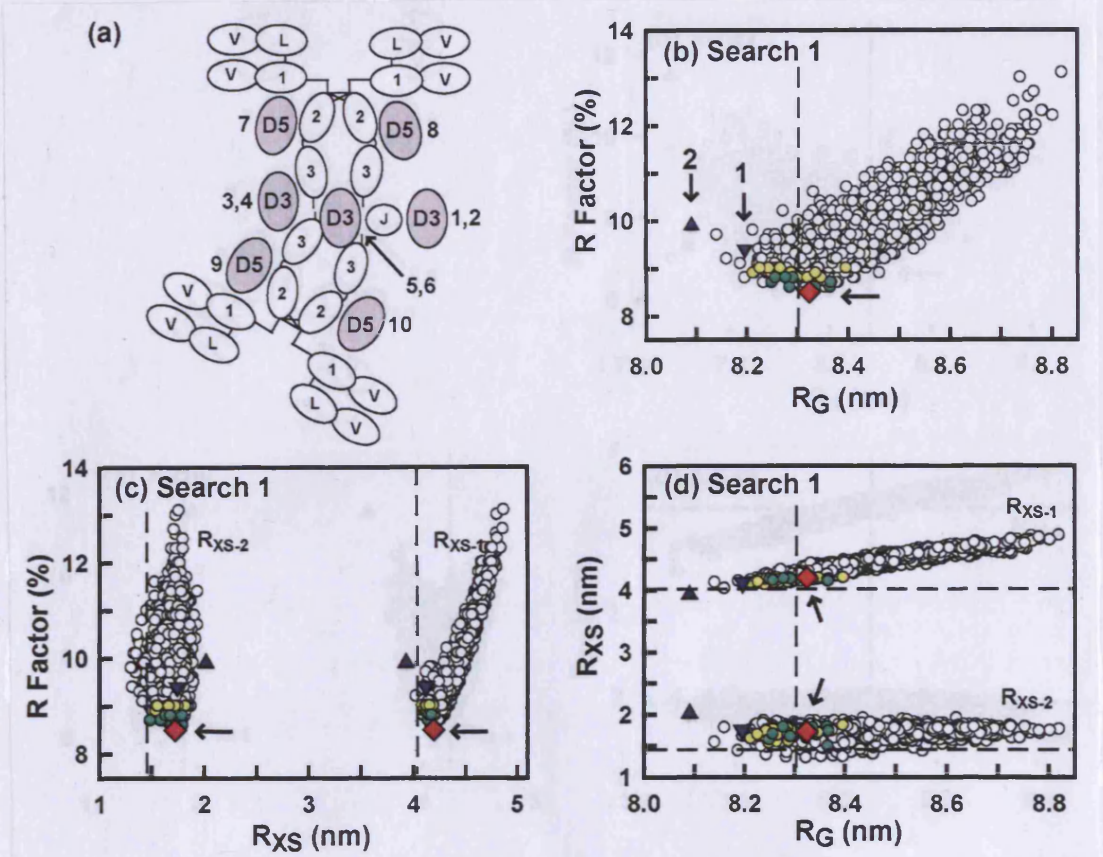

Figure 6.6 Constrained modelling Search 1 analysis of SIgA1

(a) Searches 1 to 10 are denoted by shaded ovals labelled D3 or D5 to correspond to the fixed location of the SC domain relative to dIgA1 in each search. The ovals are shown in relation to Figure 6.1, and the searches are numbered from 1 to 10. (b,c,d) The R_G, R_{XS-1} and R_{XS-2} and R factor values calculated from the 1500 models in Search 1 are compared (Table 6.2). The dashed lines correspond to the experimental R_G, R_{XS-1} and R_{XS-2} values. The 34 best-fit models (●) and the ten best-fit models (●) and of these the best-fit SIgA1 model (arrowed; ◆) are shown. Two poor-fit models 1 (▼) and 2 (▲) are shown.

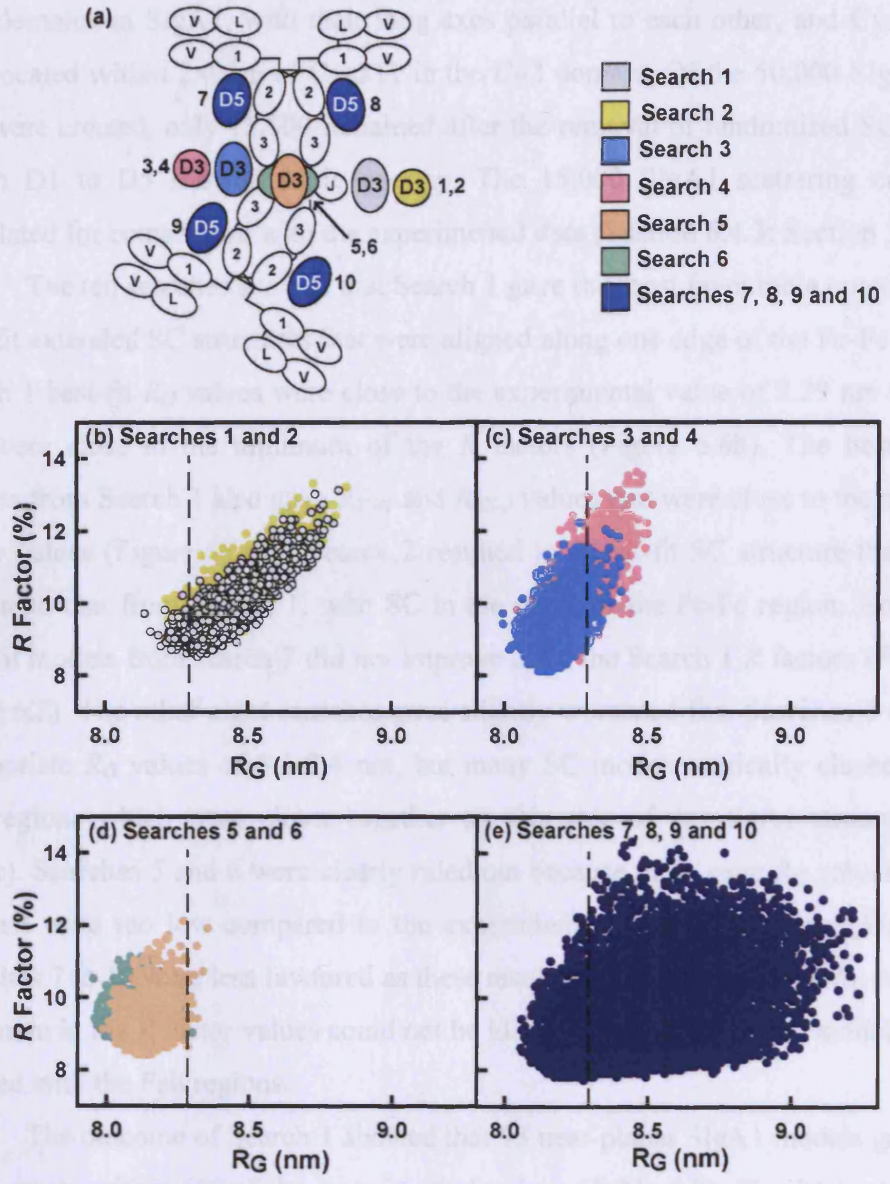

Figure 6.7 Analyses of all ten searches from the first-round of the SIgA1 constrained modelling

(a) The fixed location of the SC D3 or D5 domain relative to dIgA1 is shown for each search labelled 1 to 10 (as in Figure 6.6a). (b-e) Comparison of the X-ray R_G and R factor values for the 15,000 SIgA1 models from all ten searches. The axes remain constant for comparison. The experimental SIgA1 R_G value is shown by a dashed line. (a) Search 1 (white; Figure 6.6b), Search 2 (yellow), (c) Search 3 (cyan), Search 4 (pink), (d) Search 5 (orange), Search 6 (green) and (e) searches 7, 8, 9 and 10 (dark blue). The searches are coloured as in (a).

C_H2 domains in SIgA1, with their long axes parallel to each other, and Cys502 of D5 was located within 2.0 nm of Cys311 in the C_H2 domain. Of the 50,000 SIgA1 models that were created, only 15,000 remained after the removal of randomised SC models in which D1 to D5 showed steric overlap. The 15,000 SIgA1 scattering curves were calculated for comparison with the experimental data (Section 6.4.3; Section 3.3.4).

The ten searches showed that Search 1 gave the most favourable outcome, giving best-fit extended SC structures that were aligned along one edge of the Fc-Fc plane. The search 1 best-fit R_G values were close to the experimental value of 8.29 nm (Table 6.2) and were close to the minimum of the R factors (Figure 6.6b). The best-fit SIgA1 models from Search 1 also gave $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ values that were close to the minimum R factor values (Figure 6.6c,d). Search 2 resulted in a best-fit SC structure that was very similar to that from Search 1, with SC in the plane of the Fc-Fc region. However, the best-fit models from search 2 did not improve upon the Search 1 R factors (Figure 6.7b; Table 6.2). The other eight searches gave slightly worsened fits. Searches 3 and 4 gave appropriate R_G values of 8.1-8.4 nm, but many SC models sterically clashed with the Fab regions which were closer together on this side of the dIgA1 structure (Figure 6.7a,c). Searches 5 and 6 were clearly ruled out because these gave R_G values of 8.0-8.2 nm that were too low compared to the experimental value of 8.29 nm (Figure 6.7d). Searches 7 to 10 were less favoured as these resulted in slightly too high R_G values and a minimum in the R factor values could not be identified. Also, many SC models sterically clashed with the Fab regions.

The outcome of Search 1 showed that 93 near-planar SIgA1 models gave R_G and $R_{XS\text{-}1}$ values within 5% of the experimental values (Table 6.2). The 34 best-fit models from these 93 were selected from their R factor values and were checked visually for the absence of steric overlap between SC and dIgA1 (red, green and yellow in Figure 6.6). The ten best-fit models were selected from their R factor values (red and green in Figure 6.6). The best-fit SIgA1 structure has a good R factor of 8.5% (red in Figure 6.6) with a good visual X-ray curve fit in a Q range extending to 1 nm^{-1} (Figure 6.8a). The corresponding neutron curve fit also gave a good R factor of 9.3% despite a worsened signal-noise ratio (Figure 6.8b). The major features of the $P(r)$ curves, notably the M and L values, were reproduced (Figures 6.8a,b). Two poor-fit solutions from the best

Filter for best-fit models	Models	Hydrated spheres (N)	X-ray R_G (nm)	X-ray R_{XS-1} (nm)	X-ray R_{XS-2} (nm)	Neutron R_G (nm)	X-ray R Factor (%)	Neutron R Factor (%)	$s^{\circ}_{20,w}$ (S)
Search 1									
None	1500	4610-5000	8.14-8.82	4.03-4.88	1.31-1.94	7.18-7.68	8.5-13.1	6.8-10.7	n.a.
N, R_G, R_{XS-1}	93	4610-4970	8.14-8.40	4.03-4.20	1.42-1.90	7.18-7.34	8.5-9.8	8.7-10.7	n.a.
N, R_G, R_{XS-1}, R Factor	34	4639-4914	8.21-8.40	4.09-4.20	1.49-1.82	7.22-7.34	8.5-9.0	9.0-10.7	n.a.
N, R_G, R_{XS-1}, R Factor	10	4643-4852	8.25-8.37	4.14-4.20	1.49-1.79	7.24-7.32	8.5-8.8	8.6-10.0	11.2-11.5
Best-fit (Figure 6.8c)	1	4791	8.32	4.20	1.72	7.30	8.5	9.3	11.4
Fab reorientation fits									
None	3000	3358-4849	4.73-9.74	0.39-4.87	0.18-2.36	4.42-7.73	7.9-32.9	7.8-29.1	n.a.
N, R_G, R_{XS-1}	349	4661-4835	7.88-8.70	3.80-4.20	1.84-2.09	6.77-7.19	8.0-10.3	8.5-11.5	n.a.
N, R_G, R_{XS-1}, R Factor	50	4680-4797	7.97-8.46	3.82-4.19	1.42-1.92	6.82-7.07	8.0-8.4	9.2-10.8	n.a.
Experimental values	n.a.	4757	8.29 ± 0.20	3.90 ± 0.13	1.27 ± 0.03	7.22	n.a.	n.a.	10.8 (SEDFIT) 10.9 (DCDT+)
Poor-fit SIgA1									
Figure 6.8f model 1▼	1	4883	8.19	4.12	1.80	6.94	9.4	9.5	11.5
Figure 6.8f model 2▲	1	4947	8.09	3.93	2.02	6.96	10.8	10.1	11.7

Table 6.2 Summary of the X-ray and neutron modelling fits from Search 1 and the Fab reorientation search for the SgA1 solution structure

n.a. – not available

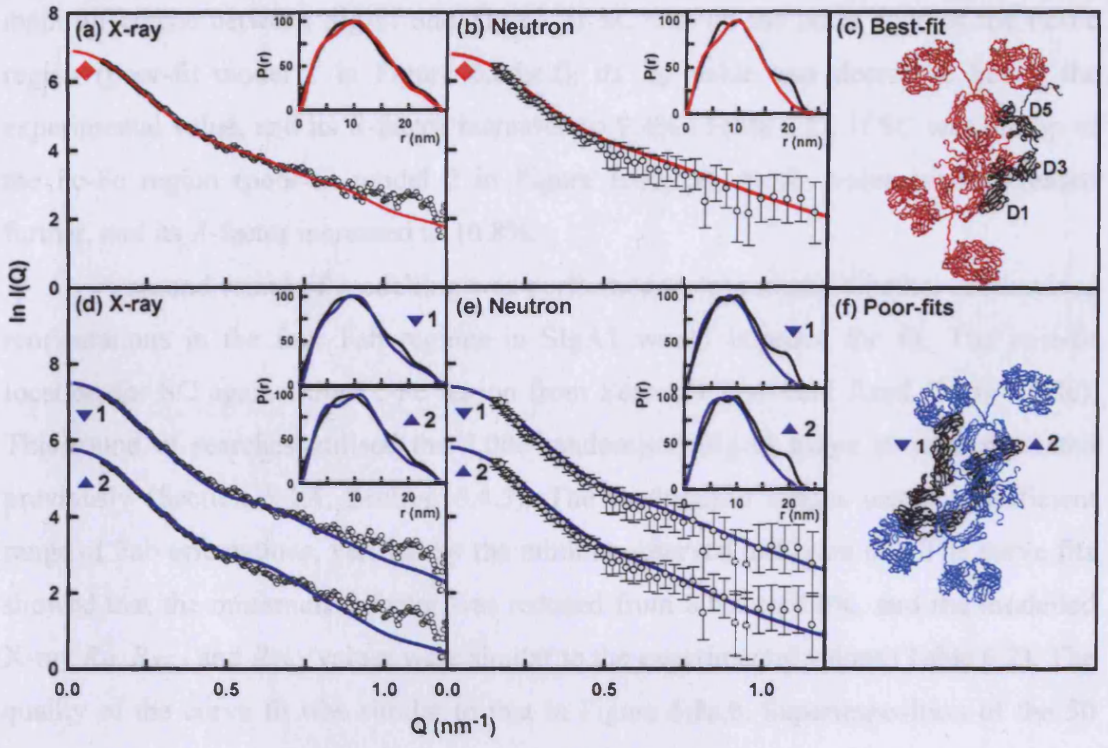

Figure 6.8 X-ray and neutron fits for the best-fit SIgA1 model from search 1 and two poor-fit SIgA1 models

(a, b) The X-ray and neutron experimental *I(Q)* and *P(r)* curves are in black, and the red lines are the modelled best-fit curves. The *P(r)* curve is shown as an inset in the top right corner. (c) The best-fit model is shown as a ribbon trace with the IgA1 dimer and J chain shown in red, and domains D1 to D5 of SC shown in black. (d, e, f) The X-ray and neutron curve fits are shown for two poor fit SIgA1 models 1 and 2. In model 1, SC is on the other edge of the Fc-Fc dimer compared to the best-fit SIgA1 model. In model 2, SC is positioned on the central face of the Fc-Fc dimer. Other details follow those in (a, b, c). The colour and shape coding of the SIgA1 models follows that in Figure 6.6.

outcome of Search 3/4 and from Search 5/6 illustrated the extent to which a best-fit structure had been determined in Search 1 (blue triangles in Figure 6.6; Figure 6.8d,e,f). Both the poor curve fits deviated in the Q range of 0.5-0.8 nm^{-1} that corresponded to the main difference between dIgA1 and SIgA1. If SC was on the other edge of the Fc-Fc region (poor-fit model 1 in Figure 6.8d,e,f), its R_G value was decreased below the experimental value, and its R-factor increased to 9.4% (Table 6.2). If SC was on top of the Fc-Fc region (poor-fit model 2 in Figure 6.8d,e,f), its R_G value was decreased further, and its R-factor increased to 10.8%.

A second round of modelling was performed to investigate whether randomised reorientations in the four Fab regions in SIgA1 would improve the fit. The best-fit location for SC against the Fc-Fc region from Search 1 was held fixed (Figure 6.8c). This round of searches utilised the 3,000 randomised dIgA1 hinge structures created previously (Section 5.4.4; Section 6.4.3). The randomised hinges tested a sufficient range of Fab orientations, verified by the minima observed in Figure 6.9. The curve fits showed that the minimum R factor was reduced from 8.5% to 8.0%, and the modelled X-ray R_G, $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ values were similar to the experimental values (Table 6.2). The quality of the curve fit was similar to that in Figure 6.8a,b. Superimposition of the 50 best-fit models showed that a limited conformational family of structures had resulted (Figure 6.10) in which the two Fab regions displayed an extended and mostly T-shaped arrangement relative to their adjacent Fc region. All four Fab regions were independent of each other. When compared with dIgA1 (Figure 5.16), the Fab regions were in approximately the same positions and included many of those revealed in the same dIgA1 fits. This agreement indicates that no conformational change in dIgA1 had occurred when this binds to SC.

The experimental sedimentation coefficient $s^0_{20,w}$ value of 10.9 S for SIgA1 was compared with the $s^0_{20,w}$ values of the ten best-fit SIgA1 models (Table 6.2). Using a previously established simulation procedure (Section 6.4.3), the calculated $s^o_{20,w}$ values for the SIgA1 models ranged between 11.2 S to 11.5 S (Table 6.2). The best-fit SIgA1 model had a $s^0_{20,w}$ values of 11.4 S. Slightly larger $s^0_{20,w}$ values of 11.5 S and 11.7 S were obtained with the two poor-fit models of Figure 6.8f. These agreements support the X-ray scattering best-fit SIgA1 model (Figure 6.8c).

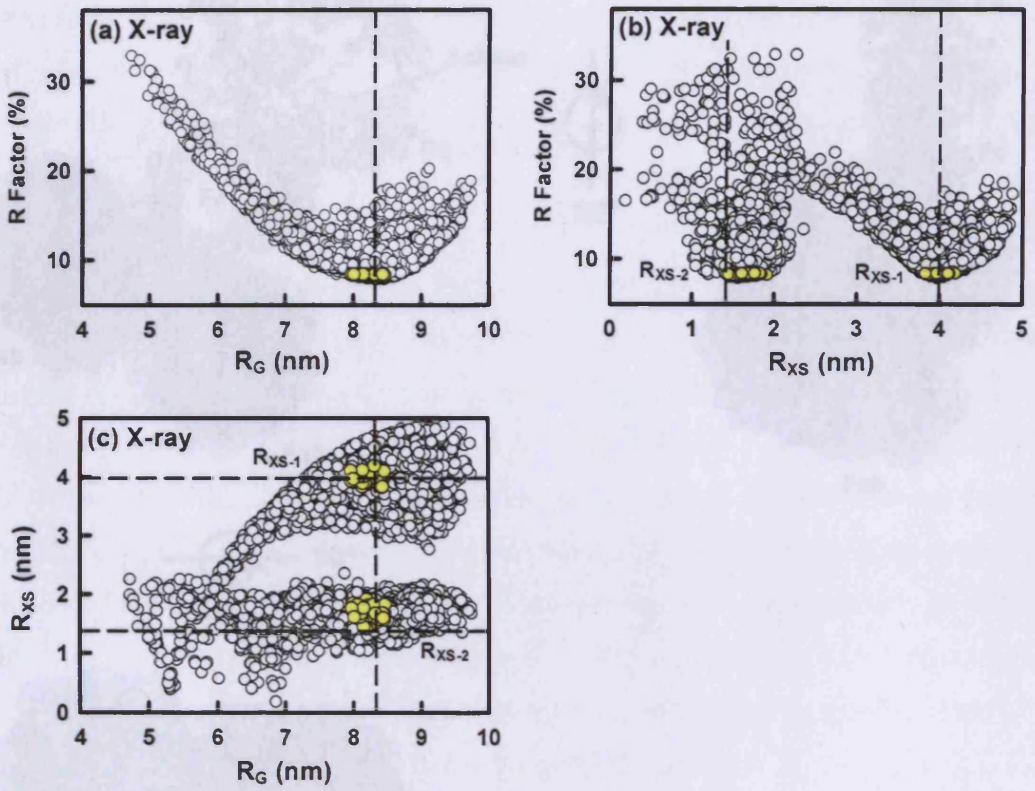

Figure 6.9 Analysis of the SIgA1 Fab reorientation second-round modelling search
Comparison of the X-ray R_G and R factor values for the 3000 models. The 50 best-fit models are shown in yellow. The vertical dashed line indicates the experimental X-ray R_G, R_{XS-1} and R_{XS-2} values.

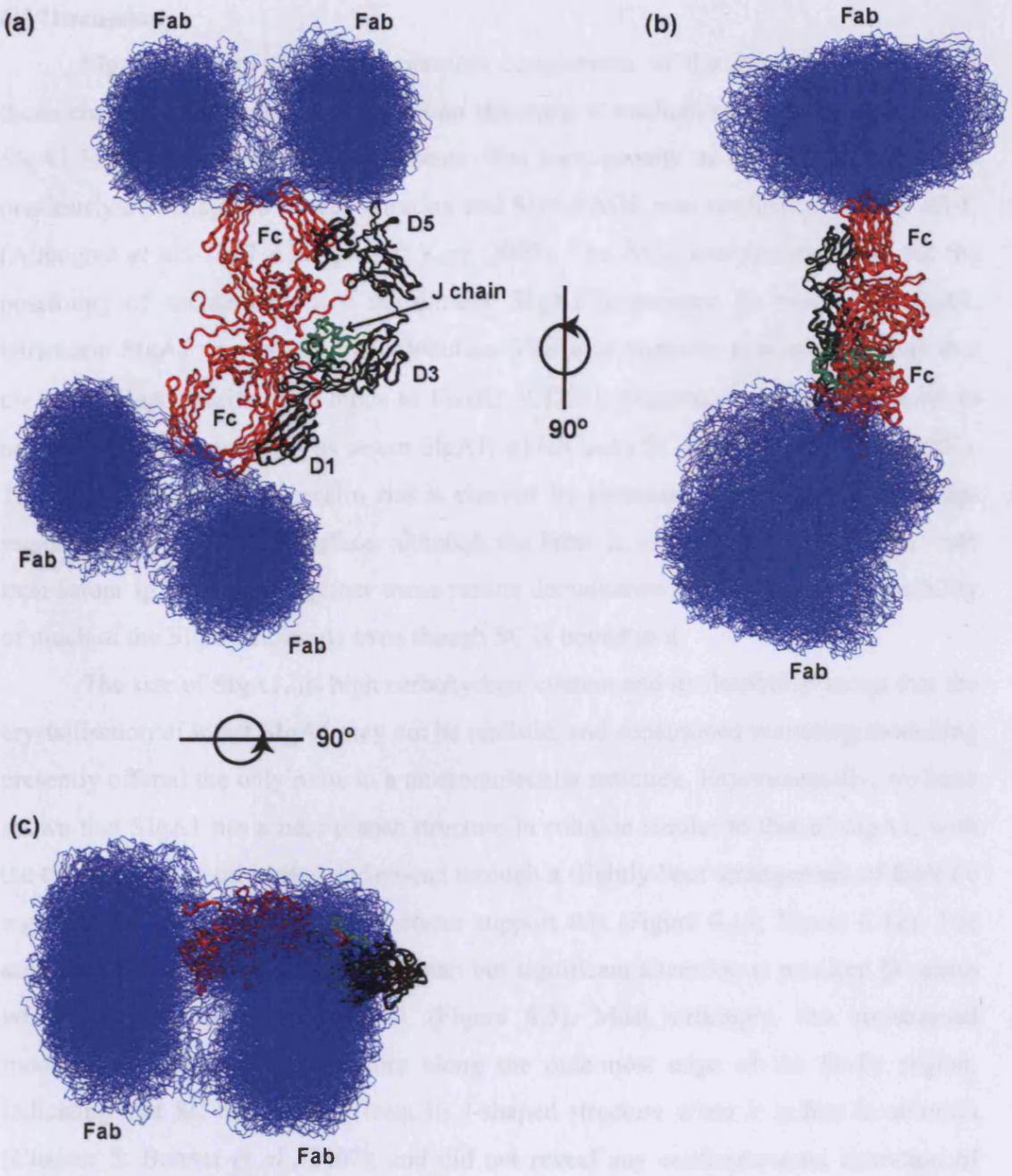

Figure 6.10 Superimposition of best 50 SIgA1 models from the Fab-refinement search

Three orthogonal views of the superimposition of the 50 best-fit SIgA1 models from the Fab reorientation second round search. The Fab regions are shown in blue, the two Fc regions are shown as red ribbons at the centre, J chain is shown in green, and SC is shown in black with the positions of D1, D3 and D5 indicated.

6.3 Discussion

SIgA is one of the most important components of the immune system. This thesis chapter presents the SIgA1 solution structure at medium resolution showing how SIgA1 is assembled from its components. The homogeneity of the SIgA1 preparation previously demonstrated by gel filtration and SDS-PAGE was confirmed here by AUC (Almogren *et al.*, 2007 Almogren & Kerr, 2007). The AUC analyses also rule out the possibility of contamination of the dimeric SIgA1 preparation by monomeric IgA1, tetrameric SIgA1 or other immunoglobulins. This also supports previous findings that the same SIgA1 purification binds to FcαR1 (CD89), triggering a respiratory burst in neutrophils in the same way as serum dIgA1, which lacks SC (Almogren & Kerr, 2007). The SIgA1 also binds to jacalin and is cleaved by proteases acting at both the hinge region and the C_H2/C_H3 interface, although the latter is cleaved at a much slower rate than serum IgA1. Taken together these results demonstrate the functional accessibility of much of the SIgA1 molecule even though SC is bound to it.

The size of SIgA1, its high carbohydrate content and its flexibility mean that the crystallisation of intact SIgA1 may not be realistic, and constrained scattering modelling presently offered the only route to a macromolecular structure. Experimentally, we have shown that SIgA1 has a near-planar structure in solution similar to that of dIgA1, with the two monomers orientated end-to-end through a slightly-bent arrangement of their Fc regions. The 34 best-fit SIgA1 structures support this (Figure 6.11; Figure 6.12). The scattering curve of SIgA1 showed a small but significant alteration at medium Q values when compared to that of dIgA1 (Figure 6.5). Most strikingly, the constrained modelling located the SC structure along the outermost edge of the Fc-Fc region, indicating that SC must unfold from its J-shaped structure when it is free in solution (Chapter 5; Bonner *et al.*, 2007), and did not reveal any conformational alteration of dIgA1 when SC was bound to it. A similar position of SC has determined for SIgA2 (Chapter 7). Because this SIgA1 structure was determined from experimental scattering data, this is not a prediction method, and the ten best-fit models qualify for deposition in the PDB. These are best viewed as a single superimposed image to appreciate the precision of the curve fit method. It is stressed that the molecular interactions of the five individual SC domains relative to the Fc-Fc region are not precisely defined by this

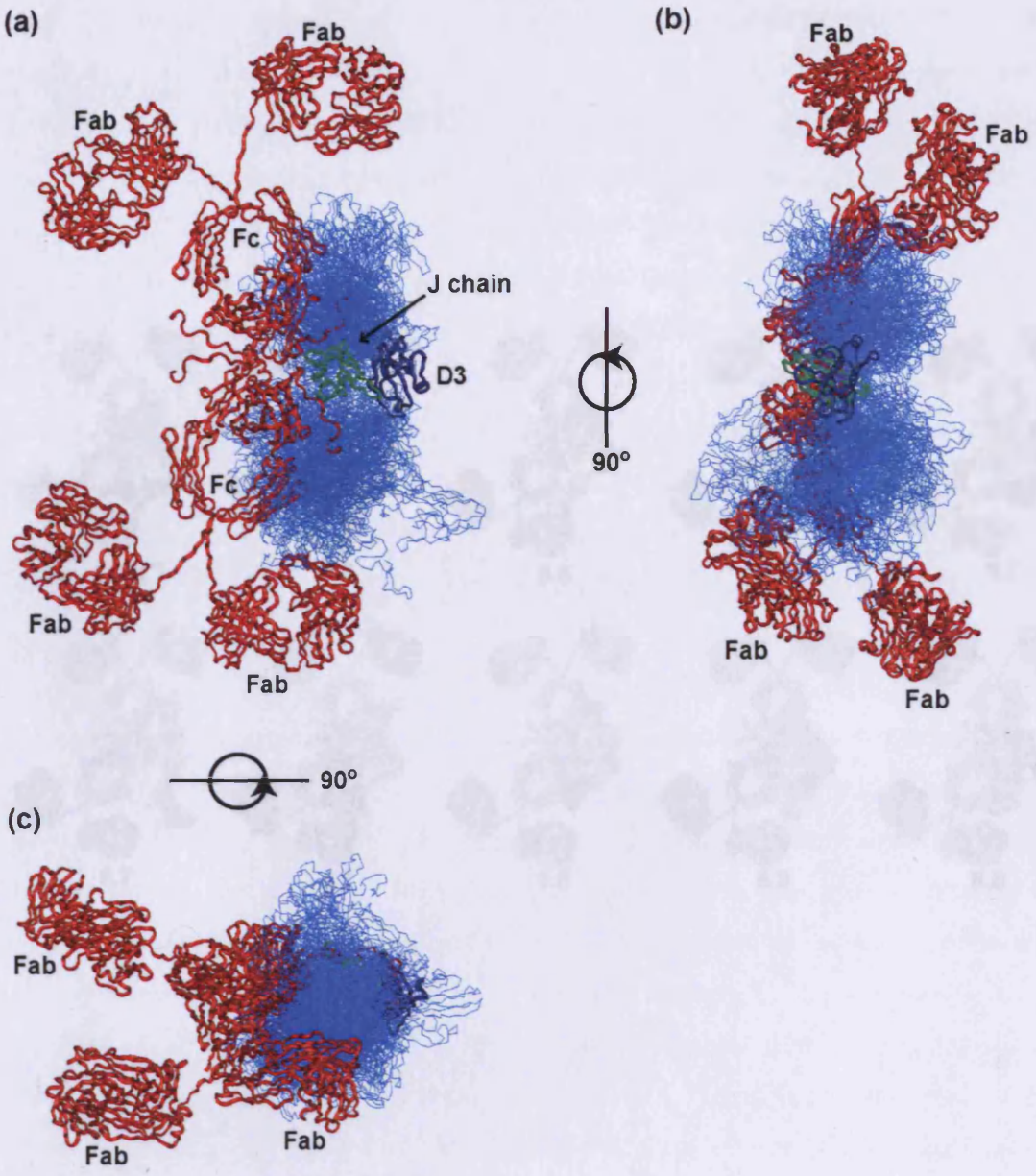

Figure 6.11 Superimposition of the 34 best-fit SIgA1 models from Search 1

The 34 best-fit models of SIgA1 from Search 1 from the first round of modelling searches (● in Figure 6.6) are shown in three orthogonal views. The two IgA1 monomers are shown in red, the J chain is shown in green and the α-carbon traces of SC are shown in blue, except for the fixed domain D3 which is shown as a ribbon.

Figure 6.12 The ten best-fit SIgA1 models from Search 1

The ten best-fit SIgA1 models are shown (● in Figure 6.6), with the dIgA1 moiety in red and blue in each one shown in the same orientation, and the variable location of SC is shown in black. The J chain is shown in green. The first model (*) corresponds to the best-fit SIgA1 model (● in Figures 6.6 and Figure 6.8). The *R* factor is shown beneath each SIgA1 model.

work. Interestingly, these best-fit SIgA1 models resemble the crystal structures for the FcαRI-Fcα, FcRn-Fcγ, FcγRIII-Fcγ and FcεRI-Fcε complexes (PDB codes 1OW0, 1A1I, 1E4K, 1F6A; Herr *et al.*, 2003b) in the sense that all four complexes show the receptor molecule making contact at a diverse range of positions on the edge of the Fc region in the three Ab classes, and not at its face. The location of SC in the best-fit SIgA1 models may also resemble the edge-to-edge contacts formed between the five Fc regions in the experimentally-determined scattering modelling structure for the human IgM pentamer (PDB code 2RCJ; Perkins *et al.*, 1991) and likewise the location of Protein A on the edge of the Fc region (PDB code 1FC2; Deisenhofer, 1981).

The strength of scattering modelling is that, while not able to determine unique molecular structures, it is able to rank the conformationally-randomised models and rule out the poorer-fit ones (Table 6.2). The exploration of three types of models in ten searches identified SC to be positioned on the most exposed, convex edge of the Fc-Fc region in dIgA1. The final models were favoured for reason of their clearly better agreement with the experimental R_G values and the avoidance of steric overlaps between the Fab regions and SC domains (Figure 6.6; Figure 6.7; Table 6.2). The original electron microscopy study of SIgA1 showed only a single Y-shaped structure, which was interpreted as two IgA1 monomers superimposed upon one another (Svehag & Bloth, 1970; Heremans, 1974; Figure 1.9). This model is ruled out by the experimental data showing that dIgA1 and SIgA1 are similarly elongated. The subsequent study indicated an end-to-end arrangement of two IgA1 monomers with various degrees of bending between them (Bloth & Svehag, 1971; Figure 1.9), and this is supported by the present scattering data. Neither of these electron microscopy studies showed the location of SC relative to dIgA1, nonetheless these SIgA images inspired many subsequent cartoons for extended SC arrangements on one face of dIgA1 (Heremans, 1974; Kerr, 1990). Many earlier hypothetical models that placed SC on one face or wrapped this around the Fc-Fc region in cartoons result in too low R_G values when modelled (e.g. Model 2 in Figure 6.8f; Table 6.2), and were not favoured by the constrained scattering modelling.

Our new SIgA1 structure offers a further understanding of the mucosal immune system. A near-planar SIgA1 structure with SC located at one edge of this would be

readily visualised to interact effectively with cell surface receptors, most notably FcαRI which is involved in IgA-mediated phagocytosis, oxidative bursts and Ab-dependent cellular cytoxicity (Section 1.6.1; Monteiro & van de Winkle, 2003). The SIgA1 model shows that receptor binding would occur independently of Ag binding as the Fc and Fab regions are well separated. Two FcαRI molecules will potentially bind at the junction between the C_H2 and C_H3 domains per Fc region, making this four per dIgA1 structure (Herr *et al.*, 2003; Pleass *et al.*, 1999). In fact, the location of SC on the outermost edge of the Fc-Fc region in SIgA1 indicates that two FcαRI sites will be blocked by SC, but the other two will remain accessible to FcαRI (Figure 6.13). The dIgA1 solution structure suggested that only two of the four potential FcαRI sites will bind to receptors on the cell surface for reason of their orientation relative to the Fc region and the cell surface (Chapter 5). Given that one FcαRI site is blocked by SC, this reasoning indicates that only a single binding site for FcαRI is expected to be available, being located on the inner edge of the Fc region (orange in Figure 6.13). This 1:1 stoichiometry is as expected (Mazengera & Kerr, 1990). Accordingly it can be predicted that SIgA1 and dIgA1 will make similar interactions with FcαRI on cell surfaces.

The new SIgA1 solution structure clarifies the formation of SIgA1 from dIgA1 and SC, and the manner in which pIgR will transcytose dIgA1 across cells. In the dIgA1 structure, the convex bend between the two Fc regions offers a potential covalently-bound location for the J chain within the crevice that is created by this bend, even though it was not possible to deduce an exact molecular location for J chain at the resolution of scattering modelling (Chapter 5). However, the positioning of the J chain is consistent with the immunologically hidden location of the J chain (Heremans, 1974). Unbound SC shows a J-shaped arrangement of domains in solution (Chapter 4; Bonner *et al.*, 2007). The new SIgA1 modelling shows that SC becomes unfolded into an extended domain arrangement along the edge of the Fc-Fc region assigned to contain the J chain. This is consistent with previous observations showing that J chain is mandatory for SC binding (Brandtzaeg & Prydz, 1984; Johansen *et al.*, 2001), and that bound and free SC can be distinguished by different conformational epitopes (Brandtzaeg, 1974). The extensive interactions between SC and the edge of the Fc-Fc region are consistent

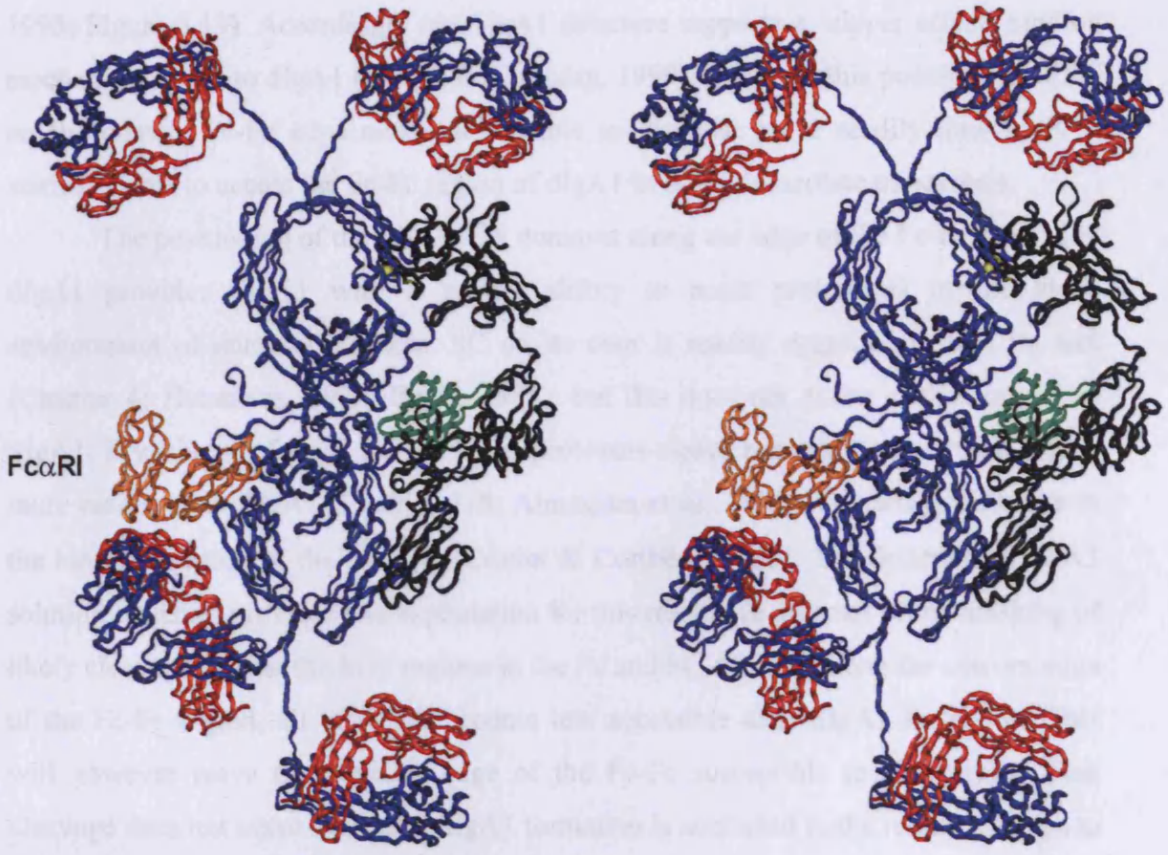

Figure 6.13 Stereo view of the best-fit SIgA1 solution structure and its interaction with an FcαR1 molecule

The best-fit SIgA1 model (Figure 6.8c; Figure 6.12) and its interaction with an FcαR1 molecule is shown as a stereo pair. The SIgA1 heavy chains (blue), light chains (red) J chain (green), SC (black) and the FcαR1 (orange; PDB code 1OW0; Herr *et al.*, 2003b) are shown. Cys311 is shown in yellow spacefill.

with biochemical studies. Thus the SC D1 domain is important for the initial interaction of SC with dIgA1, most notably the CDR regions (Figure 1.13), but D1 will not bind alone to Fc (Bakos *et al.*, 1991b; Hamburger *et al.*, 2004). The SC D2 and D3 domains increase the affinity for dIgA1 (Norderhaug *et al.*, 1999). The SC D5 domain covalently binds with Cys311 at the edge of the Cα2 domain of IgA (Fallgreen-Gebauer *et al.*, 1993; Figure 6.13). Accordingly our SIgA1 structure supports a 'zipper effect' binding mechanism of SC to dIgA1 (Crottet & Corthésy, 1999). Likewise this positioning of SC on the convex Fc-Fc edge makes it possible to visualise more readily how pIgR is sterically able to access the Fc-Fc region of dIgA1 in order to mediate trancytosis.

The positioning of the SC D1-D5 domains along the edge of the Fc-Fc regions in dIgA1 provides SIgA1 with a greater ability to resist proteolysis in the harsh environment of mucosal surfaces. SC on its own is readily degraded to half its size (Chapter 4; Heremans, 1974; Beale, 1985), but this does not occur in the context of SIgA1. Trypsin, pepsin and two bacterial proteases cleave monomeric IgA1 and dIgA1 more readily than SIgA1 (Lindh, 1975; Almogren *et al.*, 2003). SC delays cleavage in the hinge/Fc region of the α-chain (Crottet & Corthésy, 1998). The determined SIgA1 solution structure provides one explanation for this resistance in terms of the masking of likely cleavage sites at the loop regions in the Fc and SC domains along the convex edge of the Fc-Fc region, all of which become less accessible after SIgA1 formation. This will however leave the concave edge of the Fc-Fc susceptible to proteolysis. That cleavage does not occur here after SIgA1 formation is attributed to the reduced access to proteases mediated by the two Fab regions above and below it. This was suggested by visual inspection of the crystal structures for the bacterial *Pseudomonas aeruginosa* alkaline protease and the *Serratia marcescens* metalloprotease (PDB codes 1KAP and 1SRP; Baumann *et al.*, 1993; Hamada *et al.*, 1996). The formation of SIgA1 is likely to lead to a reduction of flexibility between the Fc and Fab regions, which in turn will reduce the susceptibility of surface loops to proteases at this concave edge. Accessibility to proteases at the concave edge of the Fc-Fc region will also be blocked further if SIgA1 is bound at this edge to cell surface receptors. The surface location of eight N-glycosylation sites at Asn263 and Asn459 on the dIgA1 Fc regions, togther with the seven glycosylation sites in SC when incorporated into SIgA1 will mask these surfaces

from protease cleavage (Chapter 4; Chapter 5).

The elucidation of a near-planar solution structure for SIgA1 with SC positioned on the convex edge of the Fc-Fc region has transformed our understanding of the function of SIgA. This opens the way for further experimentation in order to refine and extend our structural knowledge of SIgA assembly, and the way that SIgA1 performs its functions in mucosal immunity.

6.4 Materials and Methods

6.4.1 Preparation and composition of SIgA1

Native SIgA1 containing λ and κ class light chains was isolated from human colostrum provided by consenting healthy volunteers, and purified using a combination of thiophilic chromatography and jacalin-agarose lectin affinity chromatography (Sun *et al.*, 2005; Almogren & Kerr, 2007). These colostrum SIgA1 samples were provided by Dr. Adel Almogren and Professor Michael A. Kerr. SIgA1 samples were subjected to size-exclusion chromatography to remove non-specific aggregates and checked using reducing and non-reducing SDS-PAGE to confirm sample integrities before and after data collection. For X-ray scattering and AUC experiments, SIgA1 preparations were studied in Dulbecco's phosphate buffer saline (PBS; 12.5 mM sodium phosphate, 140 mM NaCl, pH 7.4) supplemented with EDTA and sodium azide as anti-bacterial preservatives (0.5 mM EDTA, 0.02% NaN_3). For neutron scattering, the buffer was Dulbecco's PBS as above, and made up in 100% ^2H_2O, into which the neutron samples were dialysed at 6°C for 36 h with four buffer changes. The composition of SIgA1 corresponds to two IgA1 monomers with κ light chains, J chain and SC. For the data analyses, the IgA1 and J chain amino acid and carbohydrate composition was taken from that used previously (Section 5.4.1; Boehm *et al.*, 1999), and combined with that used previously for SC (Section 4.4.1). Based on this composition, SIgA1 has a molecular mass of 424.2 kDa; an unhydrated volume of 539.0 nm^3, a hydrated volume of 712.2 nm^3 (based on a hydration of 0.3g H_2O/g of glycoprotein and an electrostricted volume of 0.0245 nm^3 per bound water molecule), an absorption coefficient at 280 nm (1%, 1 cm path length) of 12.3, and a partial specific volume $\bar{\nu}$ of 0.723 ml/g (Perkins, 1986).

6.4.2 X-ray and neutron scattering and ultracentrifugation data

SIgA1 X-ray scattering data were obtained in one beam session at the European Synchrotron Radiation Facility (ESRF), Grenoble, France on the Beamline ID02 in single-bunch mode to reduce possible radiation effects (Narayanan *et al.*, 2001; Section 2.2.3.2). The X-ray scattering data was collected by Dr. Patricia Furtado. Storage ring

currents ranged from 11 mA to 18 mA with ring energy of 6.0 GeV. The sample-to-detector distance of 3.0 m yielded a Q range of 0.08 to 2.0 nm^{-1}. SIgA1 samples were contained in water-cooled Perspex cells at 15°C, of path thickness 1 mm and mica windows of thickness 25 μm, at concentrations of 0.30 mg/ml and 0.60 mg/ml. Samples were measured in 10 time frames, each of which lasted for 0.1 sec, 1 sec or 2 sec, in order to assess and eliminate possible radiation damage effects. Neutron scattering data for SIgA1 were obtained by Dr. Patricia Furtado, in one beam session on Instrument LOQ at the pulsed neutron source ISIS at the Rutherford Appleton Laboratory, Didcot, U.K. (Heenan *et al.*, 1997; Section 2.2.3.3). The proton beam current was approximately 180 μA and data acquisition times was between 4 to 12 h for a SIgA1 concentration of 0.60 mg/ml, measured at 15°C over a Q range of 0.13 to 2.1 nm^{-1}. Other details including the data reduction procedure, the Guinier analyses and the calculation of the distance distribution function $P(r)$ using GNOM are as described previously (Section 2.2.4). For GNOM, the SIgA1 X-ray $I(Q)$ curve contained 521 data points between Q values of 0.83 to 2.00 nm^{-1} and was fitted with D_{max} set as 26 nm. The dIgA1 neutron $I(Q)$ curve contained 66 data points between Q values of 0.13 to 2.1 nm^{-1} and was fitted with D_{max} set as 24 nm.

AUC studies macromolecular structures in solution by following their sedimentation behaviour on subjecting these to a high centrifugal force. AUC analyses for SIgA1 were performed at 20°C on a Beckman XL-I instrument equipped with an AnTi50 rotor. Interference and 280 nm absorbance scans were recorded at SIgA1 concentrations of 0.20 mg/ml, 0.42 mg/ml and 0.65 mg/ml. SE data sets were each acquired over 45 h using six-sector cells with column heights of 2 mm at rotor speeds of 5,000 r.p.m., 8,000 r.p.m., 11,000 r.p.m. and 14,000 r.p.m. until equilibrium had been reached at each speed as shown by the perfect overlay of runs measured at 5 h intervals. A final run at 42,000 r.p.m. was performed at the end in order to determine the background levels for use in curve fits. Molecular weights were determined on the assumption of a single species using ORIGIN v4.1 (Microcal). The buffer density was calculated to be 1.00543 g/ml from its composition using SEDNTERP (Laue *et al.*, 1992), and the viscosity was taken as 0.01002 cp. SV data were acquired over 16 h at rotor speeds of 10,000 r.p.m., 15,000 r.p.m. and 20,000 r.p.m at concentrations of 0.15

mg/ml, 0.31 mg/ml and 0.58 mg/ml. Data were analysed using DCDT+ *g(s*)* time-derivative analyses (Philo, 2000) and SEDFIT v9.4 (Schuck, 1998) based on the continuous *c(s)* distribution model with a resolution set as 150, while the cell meniscus and bottom, the frictional ratio of 1.70, and the partial specific volume $\bar{v}$, buffer density and viscosity were held fixed, and the baseline was allowed to float.

6.4.3 Constrained modelling of SIgA1

The scattering and sedimentation data were subjected to constrained modelling in order to determine the SIgA1 solution structure (Perkins *et al.*, 2008). One constraint was the use of the determined optimised dIgA1 solution structural model (Chapter 5). This dIgA1 model was derived from 6859 models created from the monomeric IgA1 structure (PDB code 1IGA; Boehm *et al.*, 1999). This dIgA1 model was refined using a further 3000 models in which the orientation of the Fab regions were optimally positioned (Section 5.4.5). The other constraint was the use of 5,000 conformationally randomised full homology models for SC that were generated by the unrestrained linker search modelling for SC (Section 4.4.5). These were based on the crystal structure of domain D1 in SC (PDB code 1XED; Hamburger *et al.*, 2004; Figure 1.13). The J chain in the dIgA1 model was arbitrarily represented by domain D4 of SC and was located between the two Fc regions. Starting from the optimised dIgA1 structure, the first-round of searches was performed by fixing the D3 or D5 domain of the 5,000 SC models in ten distinct positions adjacent to dIgA1 (Figure 6.6a; Figure 6.7a), then evaluating the resulting SIgA1 models. Searches 1 and 2 were performed with D3 next to the J chain domain on the convex edge of the Fc region in which D3 was aligned with the plane of the Fc region or perpendicular to this. Searches 3 and 4 held D3 in two similar orientations on the other concave edge of the Fc region. Searches 5 and 6 held D3 in two similar orientations on top of the centre of the dimeric Fc region. Four searches 7 to 10 were performed with D5 next to each of the dIgA1 C_H2 domains at the top of the Fc regions.

As for dIgA1, the second-round of searches randomised the hinge between the Fab and Fc regions in the best-fit SIgA1 structure obtained from the first modelling round (Figure 6.8c; Boehm *et al.*, 1999). The SIgA1 Fab-reorientation searches were

performed by generating 3000 SIgA1 models using the six 5000-linker libraries created for the 25-residue hinge. The hinge lengths were set between 5.5 nm to 9.0 nm in 0.5 nm steps, from each of which 500 conformations were randomly selected. The hinge between the C_H1 and C_H2 domains is 220-CPVPSTPPTPSPSTPPTPSPSCCHP-244 (Boehm *et al.*, 1999; Section 5.4.4).

Each SIgA1 coordinate model was converted into Debye spheres in order to calculate X-ray and neutron scattering curves for comparison with the experimental curves. Each set of atomic coordinates for a model was placed within a three-dimensional grid of cubes. A sphere of equal volume to the cube was placed at the centre of each cube. A user-specified cutoff for the minimum number of atoms contained within a cube was determined. A cube side length of 0.531 nm in combination with a cutoff of 4 atoms consistently produced sphere models within about 1% of the total unhydrated volume of 539 nm^3 calculated from its composition. The optimal total of dry spheres of radius 0.342 nm for SIgA1 was 3600. To add the hydration shell that is visible by X-ray scattering (but not by neutron scattering), the sphere models were adapted by adding spheres to the surface of the dry models using HYPRO (Ashton *et al.*, 1997) based on 0.3 g H_2O / g glycoprotein to a total hydrated volume of 712.2 nm^3. The optimal total of hydrated spheres for SIgA1 was 4757. In order to reduce the computing time to evaluate best-fit SIgA1 structures, the 5,000 SC models were first tested for steric overlap between their five domains using a 7% volume filter to leave only 1500 models. The ten searches from the first round of modelling therefore evaluated a total of 15,000 SIgA1 models. The values of N and the modelled R_G, $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ parameters in the same Q ranges used for the experimental Guinier analyses were calculated. Models that fit within 5% of N and the X-ray R_G, $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ values were ranked using the goodness-of-fit R factor of the model. No flat-background correction was required in the neutron fits. The sedimentation coefficients $s^o_{20,w}$ were calculated from the final SIgA1 coordinate models using HYDROPRO v7c (Garcia de la Torre *et al.*, 2000). To take into account the hydration shell surrounding dIgA1, the recommended value of 0.31 nm for the atomic element radius for all atoms was used as an empirical approximation of this.

6.4.4 SIgA1 Protein Data Bank accession code

The ten best-fit SIgA1 α-carbon co-ordinate models (Figure 6.12; Table 6.2) were deposited in the Protein Data Bank with the accession code 3CHN.

Chapter Seven

Solution Structure of Human SIgA2:
Implications for Mucosal Immunity of the Non-Planar IgA2 and Near-Planar IgA1 Solution Structures in Human Secretory IgA

7.1 Introduction

IgA presents several unique features when compared with the other four human antibody classes IgG, IgM, IgD and IgE. It is the only one of the five that exists in multiple oligomeric states, and it is present as two subclasses, IgA1 and IgA2, the latter of which is found in at least two allotypic forms, IgA2m(1) and IgA2m(2). IgA exists in different forms in the different mucosal environments (Kerr, 1990). SIgA is predominantly dimeric (with some higher oligomeric forms) and slightly more SIgA1 is generally present than SIgA2 except in the colonic mucosa where there is more SIgA2 (Kerr, 1990). The major interest in IgA2 is that this is more similar to the single IgA class present in other mammalian species. IgA1 homologues have only been found in higher apes (Sumiyama *et al.*, 2002).

IgA1 and IgA2 differ the most in the hinge between the Fab and Fc regions. The IgA2 hinge is 13 residues shorter than that in IgA1 and consequently lacks the five O-glycosylation sites present in IgA1 (Torano *et al.*, 1978; Putnam *et al.*, 1979; Figure 7.1; Figure 7.2). As well as the N-linked glycosylation sites at Asn263 and Asn459 present in IgA1, the IgA2 allotypes contain two extra N-linked glycosylation sites at Asn166 and Asn337 in the C_H1 and C_H2 domains respectively (Figure 7.1; Figure 7.2). In the most common IgA2 allotype, IgA2m(1), the light chains are covalently linked by their C-terminal Cys214 residues, rather than to the heavy chains (Figure 7.1; Grey *et al.*, 1968). Like SIgA1, SIgA2 is composed of dimeric IgA2 (dIgA2) and secretory component (SC) (Figure 7.1). The two 12-domain IgA2 monomers are bound covalently through the J chain at the base of the two Fc regions (Figure 7.1). Dimeric IgA1 (dIgA1) and SC confer mutually protective properties on each other in that, alone they are susceptible to proteolysis, but when SIgA1 is formed, its resistance to proteolysis is significantly improved (Crotett & Corthésy, 1998; Lindh, 1975; Almogren *et al.*, 2003). This protection is reduced in SIgA2 (Almogren *et al.*, 2003). SC is covalently bound to a C_H2 domain in one IgA1 monomer in SIgA1, yet this covalent bond appears to be absent in SIgA2 (Fallgreen-Gebauer, 1993; Almogren *et al.*, 2007).

The solution structure of monomeric IgA2 has been determined (Furtado *et al.*, 2004; Figure 1.8b). The solution structures for dIgA1 and SC in this thesis have led to an

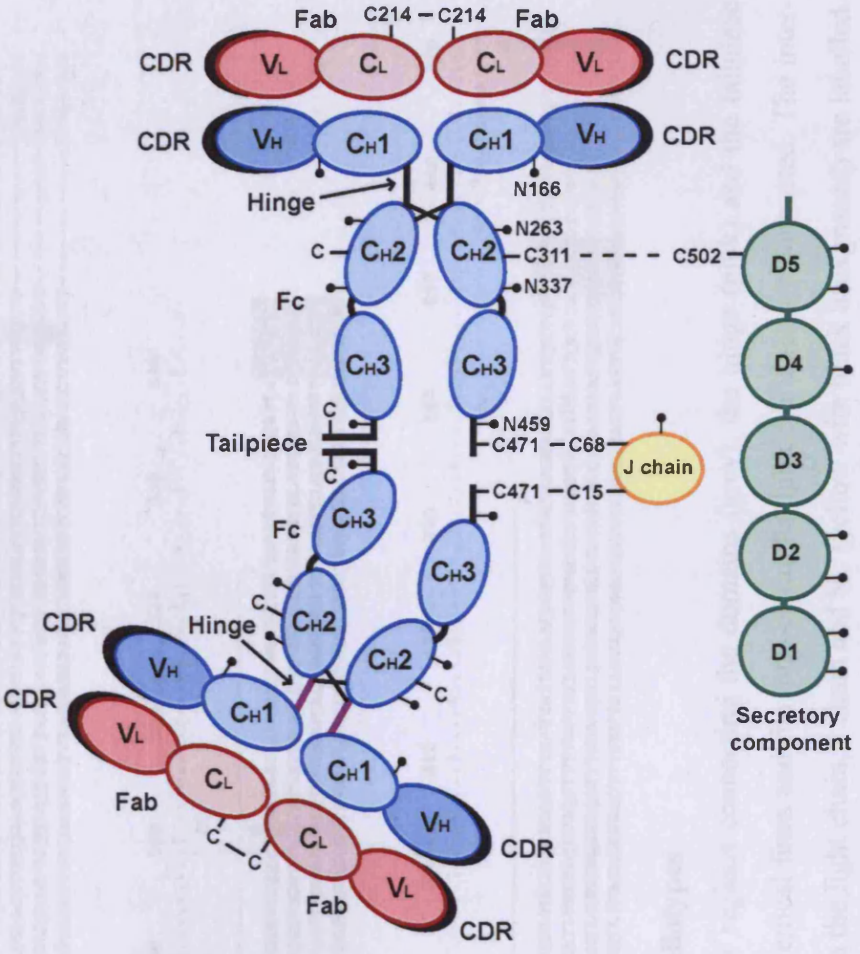

Figure 7.1 Domain structure of SIgA2

SIgA2 is composed of two IgA2 monomers, the J chain (orange outline) and SC (D1 to D5; green outline). The IgA2m(1) allotype has a Cys214-Cys214 disulphide bridge as shown, which is absent in the IgA2m(2) allotype. Each IgA2 heavy chain (blue) contains the V_H, C_H1, C_H2 and C_H3 domains, and each light chain (red) contains the V_L and C_L domains. The variable (V) and constant (C) domains are highlighted by dark and light backgrounds respectively. The CDRs (black crescents), the 13-residue hinge and the 18-residue C-terminal tailpiece (black lines) and the N-linked oligosaccharide sites at Asn166, Asn263, Asn337 and Asn459 (●) are highlighted. The inter-heavy chain disulphide bridges are denoted by an X. Cys471 in one tailpiece of each Fc region is disulphide-bridged with either Cys14 or Cys68 in the J-chain, which is located on the convex edge of the Fc-Fc region. A potential Cys311-Cys502 disulphide bridge between the C_H2 and SC D5 domain is shown as a dashed line.

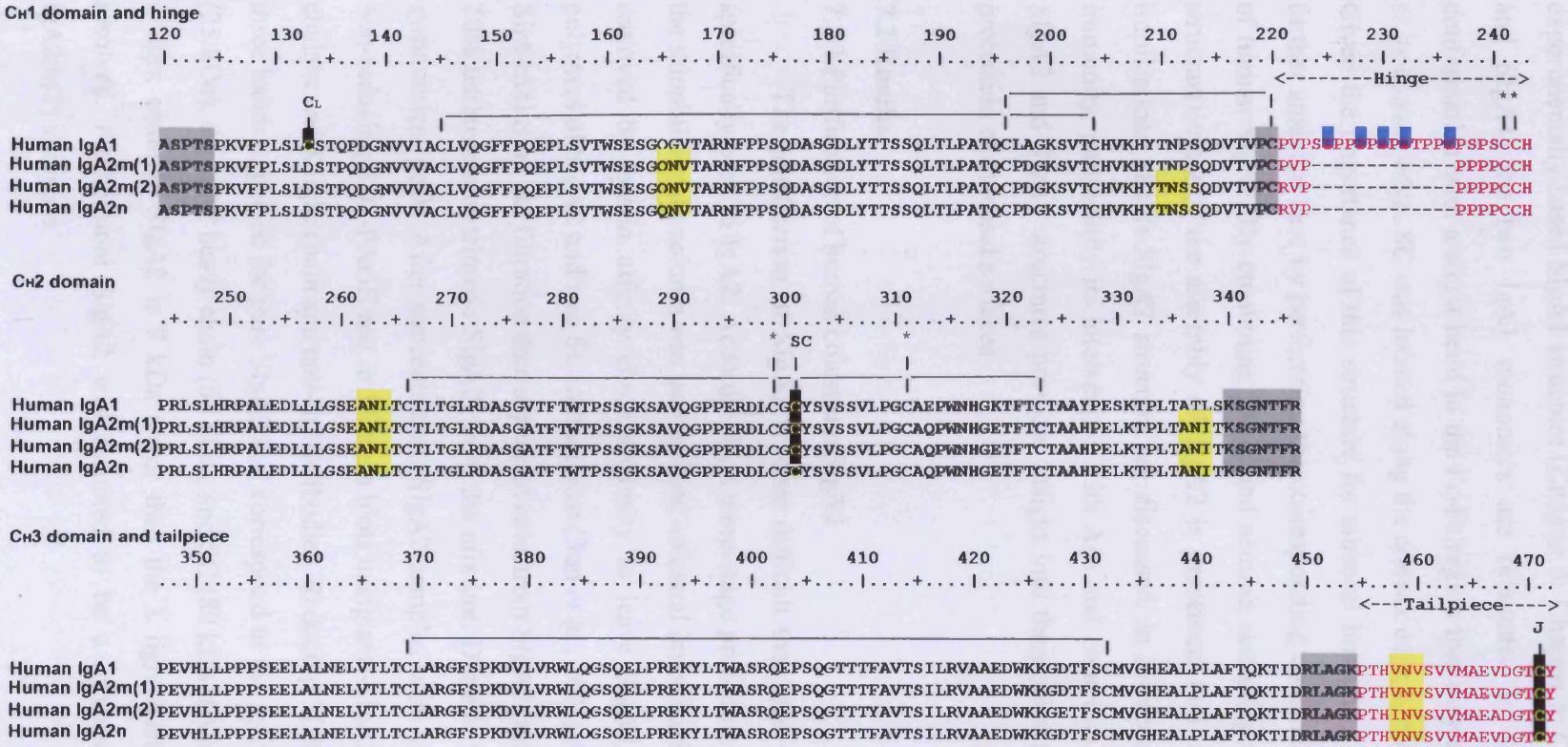

Figure 7.2 Sequence alignment of IgA1 and the IgA2 allotypes

The sequence of the C_H1, C_H2 and C_H3 domains, linker regions connecting the domains (grey), the hinge (pink) and the tailpiece (pink) are shown. Cysteine residues are highlighted by vertical lines and the intra-chain disulphide bridges are connected. The inter-heavy-chain (*) disulphide bonds and those that bind with the light chain, J chain and SC (yellow with black background) are labelled. The N-linked (yellow) and O-linked (blue) glycosylation sites are shown. The residue numbering follows that of IgA1 Bur.

experimentally-based SIgA1 structure (Chapter 4; Chapter 5; Chapter 6). In both dIgA1 and SIgA1, the two IgA1 monomers are orientated in a near-planar, tail-to-tail conformation with a slight bend in the Fc-Fc region that may accommodate the J chain at its convex edge. SC was located along the convex edge of the Fc-Fc region in SIgA1. Given the importance of this structure for mucosal immunity, the structure of SIgA is further analysed here by performing the corresponding constrained scattering modelling of human SIgA2. By combining X-ray and neutron scattering with AUC data, the first structural view of the assembly of SIgA2 is presented, and compared to SIgA1. The implications of this SIgA2 structure are discussed, in reference to its role in mucosal immunity, particularly its interactions with Ags and receptors. The comparison of the SIgA2 and SIgA1 structures provides insight into their roles that lead to the efficient protection of mucosal surfaces.

7.2 Results

7.2.1 Purification of human colostrum SIgA2

The purification of SIgA2 is more difficult than that for SIgA1 as no lectins specifically bind to IgA2. Accordingly, a three-stage procedure was employed in which the thiophilic resin column was used to bind colostral immunoglobulins, then SIgA1 was removed by jacalin affinity chromatography to leave SIgA2, secretory IgM, higher polymers of SIgA2 and free SC in the eluate (Sun *et al.*, 2005; Almogren & Kerr, 2008). Size-exclusion gel filtration then separated colostrum SIgA2 (mainly dimeric, with about 10% tetrameric or trimeric SIgA2) from the mixture. Dimeric SIgA2 eluted as a single symmetrical peak. After concentration, SIgA2 samples were checked by reducing and non-reducing SDS-PAGE and in Western blots using anti-SIgA2 α-chain, anti-SIgA2 α-chain κ and λ light chain and anti-SC antibodies. Reducing SDS-PAGE showed only the three bands expected for pure SIgA2 that correspond to the IgA2 light chain and J-chain (25 kDa), the IgA2 heavy chain (60 kDa), and SC (80 kDa) (Almogren *et al.*, 2007). The κ light chain in SIgA2 is 7 kDa larger than the λ light chain, hence this cannot be resolved. Furthermore SIgA2 was assumed to be a mixture of the IgA2m(1) and IgA2m(2) allotypes.

7.2.2 X-ray and neutron scattering of SIgA2

Comparison of the X-ray scattering *I(Q)* data from a single time frame and those averaged over ten consecutive frames showed no radiation damage or aggregation effects, and the merged ten time-frames were used for analyses for reason of improved signal-noise ratios. At the lowest Q values ($Q = 4\ \pi \sin\theta / \lambda$; 2θ = scattering angle; λ = wavelength), the Guinier analyses resulted in linear plots, from which the R_G value was obtained within satisfactory $Q.R_G$ limits (Figure 7.3a,b). IgA2 myelomas are extremely rare, accordingly no scattering comparison could be made for dIgA2, unlike that between SIgA1 and dIgA1 (Table 7.1). The mean SIgA2 X-ray R_G from five values was 8.13 ± 0.1 nm (Figure 7.3a;), which is slightly less than that of 8.29 ± 0.20 nm for SIgA1 (Table 7.1). The SIgA2 anisotropy ratio R_G/R_O (where R_O is the R_G value of the sphere with the same volume as the glycoprotein) was 1.90, which is slightly less than that of 1.93 for SIgA1 (Table 7.1). The same trend but with a larger difference was seen for monomeric IgA2m(1) and IgA1 for which the R_G/R_O ratios were 1.66 and 1.99 respectively (Table 7.1). The similarities between the two SIgA isoforms suggest that they both have a similar extended arrangement of IgA monomers with SC located at its centre (Figure 7.1). The difference between the two isoforms is attributed to the variation in the length of their hinge regions, where that for IgA2m(1) is 10 residues while that for IgA1 is 23.

The shorter axes of SIgA2 were monitored using the $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ parameters from X-ray cross-sectional Guinier analyses. The *ln(I(Q).Q)* plots revealed two linear regions in Q ranges between 0.20 – 0.28 nm^{-1} and 0.72 – 1.04 nm^{-1} that were similar to those for SIgA1 and dIgA1 (Section 5.2.2; Section 6.2.2). The mean $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ values were 4.22 ± 0.09 nm (5 values) and 1.93 ± 0.03 nm (4 values) respectively (Figure 7.3c,d; Table 7.1). In comparison, the $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ values for SIgA1 were slightly lower at 3.90 ± 0.13 nm and 1.27 ± 0.03 nm respectively, and the same was seen for dIgA1 (Table 7.1). The $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ values for monomeric IgA2m(1) and IgA1 are similar at 2.47 nm and 1.47 nm respectively, and 2.20 nm and 1.56 nm respectively, however the $R_{XS\text{-}1}$ values are lower than those of dIgA1 (Table 7.1). The larger $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ values for SIgA2 may result from an alteration in the relative arrangement of the

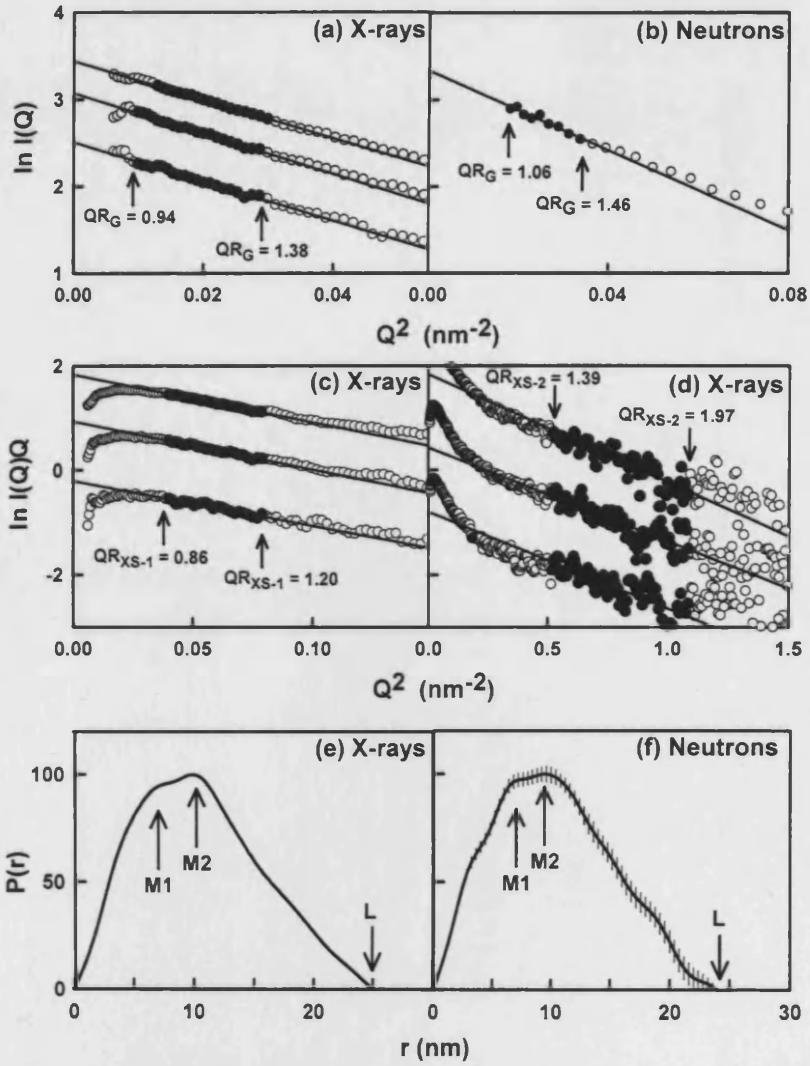

Figure 7.3 Guinier plots and distance distribution functions *P(r)* analyses for SIgA2

(a-d) The $Q.R_G$ and $Q.R_{XS}$ ranges used to determine the R_G, $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ values are represented by filled circles between the arrowed data points. The X-ray fits correspond to concentrations of 0.90 mg/ml (upper), 0.60 mg/ml (middle), and 0.32 mg/ml (lower). The neutron fit corresponds to a concentration of 1.0 mg/ml.(a,b) The X-ray and neutron Q ranges used for the R_G values were 0.10 to 0.17 nm^{-1} and 0.10 to 0.18 nm^{-1} respectively and (c,d) those for the $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ values were 0.20 to 0.28 nm^{-1} and 0.72-1.04 nm^{-1} respectively. (e,f) The two most frequently occurring distances in the *P(r)* curve *M*1 and *M*2 are 7.3 nm and 10.0 nm respectively (X-ray) and 7.2 nm and 9.6 nm (neutron) respectively. The SIgA2 maximum length *L* is 25 nm (X-ray) and 24 nm (neutron) when *P(r)* reaches 0. Error bars are shown for the SIgA2 neutron *P(r)* plot.

Protein	Data	Guinier analyses				*P(r)* analyses			
		R_G (nm)	R_{XS-1} (nm)	R_{XS-2} (nm)	R_G/R_O	R_G (nm)	L (nm)	$M1$ (nm)	$M2$ (nm)
Secretory IgA2	X-ray	8.13 ± 0.10	4.22 ± 0.09	1.93 ± 0.03	1.90	8.19 ± 0.37	25.0	7.3	10.0
	Neutrons	7.57	n.a.	n.a.	1.77	7.95	24.0	7.2	9.6
Secretory IgA1	X-ray	8.29 ± 0.20	3.90 ± 0.03	1.27 ± 0.03	1.93	8.30 ± 0.17	26.0	7.0	9.6
(Chapter 6)	Neutrons	8.27± 0.33	n.a.	n.a.	1.85	8.04	24.0	6.8	10.0
Dimeric IgA1	X-ray	8.65 ± 0.27	3.94 ± 0.18	1.43 ± 0.07	2.16	8.67 ± 0.17	26.0	4.9	9.9
(Chapter 5; Bonner *et al.*, 2008)	Neutrons	7.60 ± 0.05	n.a.	n.a.	2.08	7.47	23.0	5.1	10.1
Monomeric IgA2m(1)	X-ray	5.18 ± 0.09	2.47 ± 0.09	1.47 ± 0.08	1.66	5.21 ± 0.15	17.0	4.5 ± 0.3	6.8 ± 0.7
(Furtado *et al.*, 2004)	Neutrons	5.03 ± 0.01	2.21 ± 0.10	1.04 ± 0.06	n.a.	5.12 ± 0.01	17.0	4.5	n.a.
Monomeric IgA1	X-ray	6.20 ± 0.13	2.20 ± 0.26	1.56 ± 0.16	1.99	6.12	21.0	3.7	9.1
(Boehm *et al.*, 1999)	Neutrons	6.11 ± 0.18	2.17 ± 0.23	1.18 ± 0.12	n.a.	n.a.	n.a.	n.a.	n.a.
Monomeric IgA1-HSA	X-ray	7.47 ± 0.31	2.31 ± 0.11	1.31 ± 0.05	2.13	7.36 ± 0.01	25.0	4.87	8.25
(Almogren *et al* 2006)	Neutrons	7.09 ± 0.01	n.a.	n.a.	2.22	n.a.	n.a.	n.a.	n.a.
Secretory component	X-ray	3.53 ± 0.43	1.76 ± 0.08	n.a.	1.47	3.66 ± 0.19	13.0	3.8	n.a.
(Chapter 4; Bonner *et al.*, 2007)	Neutrons	3.63 ± 0.28	1.30 ± 0.10	n.a.	n.a.	3.73 ± 0.17	12.0	3.2	n.a.

Table 7.1 Comparison of X-ray and neutron data analyses for the polymeric forms of human IgA and its secretory component

All proteins are from human sources except for SC which is recombinant and IgA2m(1) which is a recombinant human-mouse chimaera. The X-ray data were obtained on instrument ID02 at ESRF except for that for monomeric IgA1, which was from Stations 2.1 and 8.2 at the SRS Daresbury, and the neutron data were all obtained on Instrument LOQ at ISIS. n.a. – not available

Fab and Fc regions within the dimeric structure when compared to the near-planar SIgA1 and dIgA1 structures.

Neutron scattering data acts as a control against possible X-ray-induced radiation damage, any internal structural inhomogeneity caused by the 13% carbohydrate content of SIgA2, and any hydration effect which is less visible by neutrons but not by X-rays (Perkins *et al.*, 2008; Perkins, 2001). SIgA2 data were collected using a ^2H_2O buffer to correspond to a high negative glycoprotein-solvent contrast, in distinction to the high positive contrast observed by X-rays. The SIgA2 neutron R_G value was 7.57 nm (Figure 7.3b; Table 7.1), this being 0.56 nm lower than the X-ray value, for which the difference is attributed to a largely invisible hydration shell (Perkins, 2001). The neutron anisotropy ratio R_G/R_O was 1.77, and is slightly less than the X-ray value of 1.90. This ratio of 1.77 is less than those of 1.85 and 2.08 for SIgA1 and dIgA1, which is consistent with the relative values of the three X-ray ratios (Table 1.7). No neutron $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ values could be calculated for reason of an insufficient signal-noise ratio. The Guinier *I(0)/c* values for proteins measured in ^2H_2O buffers and normalised against a standard deuterated polymer leads to the protein molecular mass M (where $M = I(0)/c \times 9.10^5$) (Boehm *et al.*, 1999). The neutron Guinier *I(0)/c* value of 0.47 resulted in a mass of 425 Da for SIgA2 (Figure 2.5), in good agreement with the composition-derived value of 424 kDa for SIgA2, showing that the scattering data corresponded to the expected composition of SIgA2, which included four κ light chains and 24 N-linked biantennary oligosaccharides (Figure 1). The replacement of κ light chains by λ light chains in SIgA2 will result in a decrease of 7 kDa in mass.

The calculation of the distance distribution function *P(r)* from the *I(Q)* curve provides structural information in real space. The mean R_G values determined from the X-ray and neutron *P(r)* curves were 8.19 ± 0.37 nm (4 values) and 7.95 nm respectively (Figure 7.3e,f), in good agreement with the Guinier analyses (Table 7.1). The SIgA2 *P(r)* curves reproducibly showed two broad maxima *M1* and *M2* located at 7.3 nm and 10.0 nm respectively (X-ray) and 7.2 nm and 9.6 nm respectively (neutron). While these were also observed for SIgA1 and dIgA1, the *M1* values for dIgA1 were slightly smaller than those for SIgA1 and SIgA2 (Table 7.1). The comparable appearance of the SIgA2 and

SIgA1 *P(r)* curves indicates similarity in their overall solution structures, while the change in *M1* for dIgA1 is attributed to the absence of SC in this. The maximum length, *L*, is determined from where the *P(r)* curve reaches zero at larger *r* values. For SIgA2, *L* was determined to be 25 nm (X-ray) and 24 nm (neutron) (Figure 7.3e,f). The *L* values for SIgA1 and dIgA1 (Table 7.1) were similar to those for SIgA2 and provide further evidence to show that SIgA2 possesses a similar extended topology to that for SIgA1 and dIgA1.

7.2.3 Analytical ultracentrifugation of SIgA2

In order to confirm the molecular mass of SIgA2, sedimentation equilibrium (SE) experiments were performed at three concentrations and four rotor speeds (Section 7.4.2). Assuming that a single species was present, SE analyses determined molecular mass values for SIgA2 that ranged between 320 kDa to 490 kDa (Figure 7.4). Their extrapolation to zero concentration gave a mean molecular mass of 415 ± 60 kDa, which is within error of the sequence-derived value of 425 kDa. This confirmed the neutron-derived molecular mass calculation (Section 7.2.2).

The SIgA2 sedimentation coefficient $s^{o}_{20,w}$ was determined from sedimentation velocity (SV) experiments. SEDFIT analyses showed a good visual agreement between the experimental and fitted boundaries (Figures 7.5a,c). The resulting *c(s)* distribution analyses revealed that a single SIgA2 species was present with an $s^{\circ}_{20,w}$ value of 12.2 S and 12.1 S from the interference and absorbance data respectively (Figures 7.5b,d). No other species were detectable and this confirmed the monodispersity of the SIgA2 preparation. The conversion of the *c(s)* plots to *c(M)* mass distributions showed that the 12.2 S species corresponded to a molecular mass of 422 kDa (data not shown), which is consistent with the neutron, SE and sequence-derived molecular mass values. The DCDT+ *g(s*)* time derivative analyses confirmed the *c(s)* analyses. Between 6-14 scans at the centre of the sedimentation profile were analysed to give an $s^{\circ}_{20,w}$ value of 12.1 S for both interference and absorbance data sets (Figure 7.5e,f). The SIgA2 sedimentation coefficient of 12.1 S is larger than that of SIgA1 at 10.8 S. The frictional ratio f/f_o (where f_o is the frictional coefficient of a sphere with a volume equal to that of the hydrated

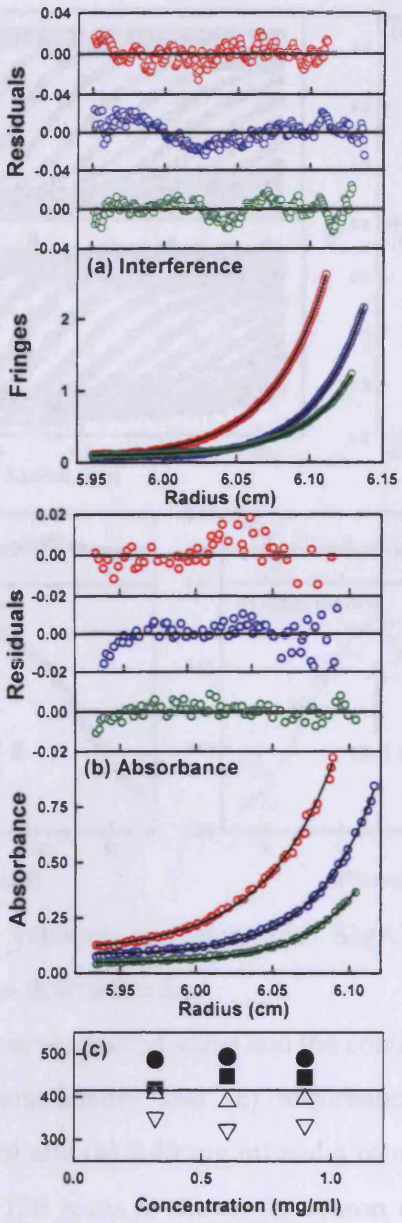

Figure 7.4 Sedimentation equilibrium analyses for SIgA2

SIgA2 SE analyses are shown using (a) interference and (b) absorbance optics at concentrations of 0.90 mg/ml (red), 0.60 mg/ml (blue), and 0.32 mg/ml (green) at a rotor speed of 8,000 r.p.m. Circles represent the experimental data and the continuous black lines represent their fits. The corresponding curve fit residuals are shown above the exponential fits. (c) The rotor speed dependence of the fitted molecular mass gave a mean value of 415 ± 4 kDa denoted by (●) 7,000 r.p.m, (■) 9,000 r.p.m, (△) 11,000 r.p.m and (▽) 14,000 r.p.m.

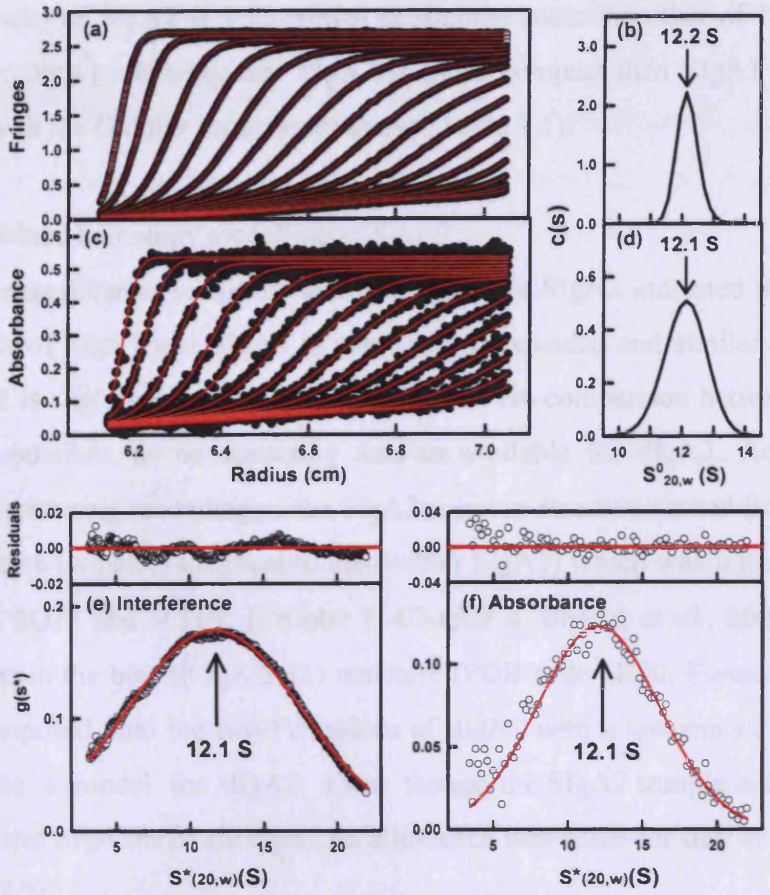

Figure 7.5 Sedimentation velocity analyses for SIgA2 using SEDFIT *c(s)* distribution and DCDT+ time-derivative fits

The black circles represent the experimental scans and the continuous red lines represent the fits. (a,c) SEDFIT (a) interference and (c) absorbance analyses at a SIgA2 concentration of (a) 0.93 mg/ml and (b) 0.40 mg.ml and a rotor speed of 20,000 r.p.m., where every fifth scan of the 120 scans is shown for reason of clarity. The 120 scans were recorded at 5 min intervals between consecutive ones. The corresponding *c(s)* plots from (a,c) are shown from which the sedimentation coefficient for SIgA2 was determined to be 12.2 S (interference) and 12.1 (absorbance). (e,f) DCDT+ *g(s*)* analyses, where the goodness-of-fit residuals are shown above each analysis. The arrowed $s_{20,w}$ value was determined to be 12.1 S based on data at 0.93 mg/ml (absorbance and interference), at rotor speeds of 10,000 r.p.m. (interference) and 15,000 r.p.m. (absorbance) The fits were determined using from 4 to 14 scans recorded mid-way through the experiment shown in (a,c).

macromolecule) of SIgA2 is 1.52, which is slightly lower than that of 1.70 for SIgA1 and 1.63 for dIgA1, showing that SIgA2 is more compact than SIgA1 and dIgA1 in agreement with the Guinier R_G analyses above (Table 7.1).

7.2.4 Constrained homology modelling of SIgA2

The experimental scattering and AUC data for SIgA2 indicated that the overall arrangements of SIgA1 and SIgA2 in solution are extended and similar to each other, while SIgA2 is slightly more compact than SIgA1. No comparison between SIgA2 and dIgA2 was possible, as no scattering data are available for dIgA2. Accordingly the constrained scattering modelling of the SIgA2 solution structure started from the best-fit dIgA1 structure (which is identical to that within SIgA1) which was initially held fixed (PDB codes 2QTJ and 3CHN; (Chapter 5; Chapter 6; Bonner et *al.*, 2008). The Fc α-carbon atoms in the best-fit IgA2m(1) structure (PDB code 1R70; Furtado *et al.*, 2004) were superimposed onto the two Fc regions of dIgA1 with a low r.m.s.d value of 0.58 nm to create a model for dIgA2. Even though the SIgA2 sample comprises both the SIgA2m(1) and SIgA2m(2) allotypes, no allowance was made for this in the modelling fits (Figure 7.2).

In the first cycle of modelling searches, ten different locations of SC relative to the fixed dIgA2 structure were evaluated in models for SIgA2. This search is similar to the first round of modelling searches for SIgA1 (Section 6.4.3). The SC models were based on the 5,000 models with four conformationally-randomised linkers that were used to determine the SC solution structure by constrained scattering modelling (Section 4.4.4). Three types of searches were performed by holding one of the SC domains proximate to a specified domain in the dIgA2 structure. (i) Searches 1 to 4 explored positions for SC D3 on the edge of the planar bent dimeric Fc region in dIgA2. Hence Search 1 held the long axis of D3 parallel to the long axis of the Fc-Fc regions and next to the putative J chain position on the convex edge of the Fc-Fc regions. Search 2 repeated this with D3 now rotated by 90° to be perpendicular in position to that used in Search 1. Searches 3 and 4 repeated these searches on the other concave edge of the Fc-Fc regions. (ii) Searches 5 and 6 explored locations for SC on top of the planar Fc-Fc

regions. These positioned the long axis of D3 either parallel or perpendicular to the long axis of the Fc-Fc regions. These models correspond to several tentative models of this type in which SC is postulated to be wrapped around the planar Fc-Fc region, even though they have no experimental basis (Kerr, 1990; Mestecky *et al.*, 2005; Brandtzaeg, 2007; Heremans, 1974; Corthésy, 2003; Mestecky & McGhee, 1987; Woof & Kerr, 2006). (iii) Searches 7 to 10 explored SC models based on the presumption that in SIgA1 Cys502 in SC D5 forms a disulphide bridge with Cys311 in one of the four C_H2 domains (dashed line in Figure 7.1; Fallgreen-Gebauer *et al.*, 1993). However, recent evidence suggests this covalent bridge is absent in SIgA2 (Almogren *et al.*, 2007). Nonetheless D5 of SC was positioned next to each of the four C_H2 domains in SIgA2, with their long axes parallel to each other, and Cys502 of D5 was located within 2.0 nm of Cys311 in the C_H2 domain. Of the 50,000 SIgA2 models that were created, only 15,000 remained after the removal of randomised SC models in which D1 to D5 showed significant steric overlap. The 15,000 scattering curves were calculated for comparison with the experimental data (Section 7.4.3; Section 3.3.4).

The ten searches showed that, like the SIgA1 curve fits, Search 1 gave the most favourable outcome. This gave SIgA2 structures in which extended SC structures were aligned along the convex edge of the Fc-Fc plane. The distributions of the 1500 sets of R_G and R_{XS} values showed that the best-fit models had R_G, $R_{XS\text{-}1}$ and $R_{XS\text{-}1}$ values close to the experimental values of 8.13 nm, 4.22 nm and 1.93 nm respectively (Figure 7.6). The best-fit models from Search 1 also gave R_G, $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ values that were close to the minimum R factor values (Figures 7.6b,c). Search 2 resulted in a best-fit SC structure that was very similar to that from Search 1, with SC in the plane of the Fc-Fc region (yellow in Figures 7.7b). The other eight searches gave worsened fits (Figure 7.7). Searches 3 and 4 gave reduced R_G values of 7.5-8.1 nm, in which many SC models sterically clashed with the two Fab regions which were relatively close in the dIgA2 structure (Figure 7.7c). Searches 5 and 6 with SC located on the face of the Fc-Fc region were clearly ruled out because these gave too low R_G values of 7.5-7.9 nm (Figure 7.7d). Searches 7 to 10 gave a reasonable range of SIgA2 structures with R_G values close to the minimum of the 6,000 R factor values (Figure 7.7e). However 1760 (29%) of the 6,000

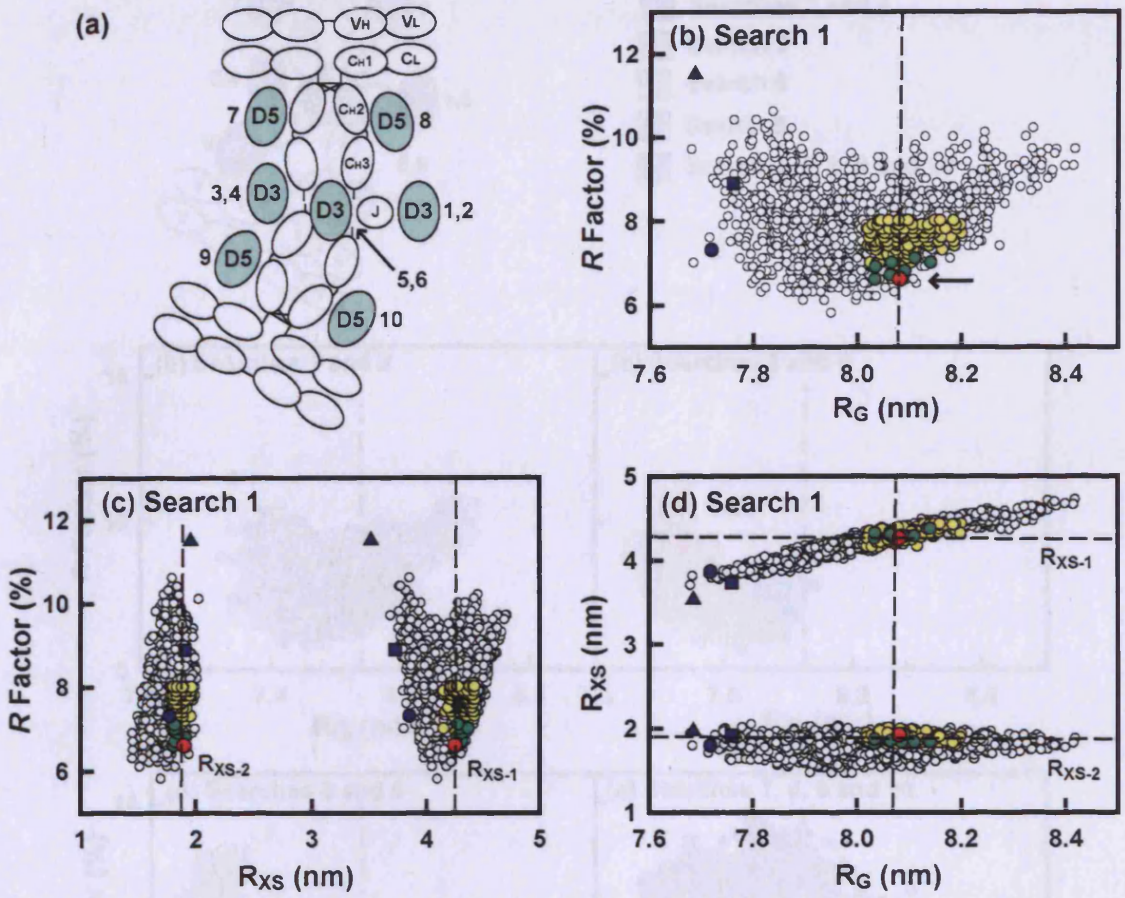

Figure 7.6 Constrained modelling Search 1 analysis of SIgA2

(a) The ten searches numbered 1 to 10 are denoted by green shaded circles labelled D3 or D5 to correspond to the SC domain held fixed relative to the two IgA2 monomers in each search. The shaded circles are shown in relation to the outline of Figure 7.1. Thus Search 1 held SC D3 fixed next to the J chain. (b,c,d) The X-ray R_G, R_{XS-1} and R_{XS-2} values calculated from the 1500 models in Search 1 are compared (Table 7.2). The 50 best-fit models (●), the ten best-fit models (●) and the best-fit model (horizontal arrow; ●) are shown. Three representative poor-fit SIgA2 models are shown (●, ■ , ▲; Table 7.2; Figure 7.8). The dashed lines correspond to the experimental R_G, R_{XS-1} and R_{XS-2} values.

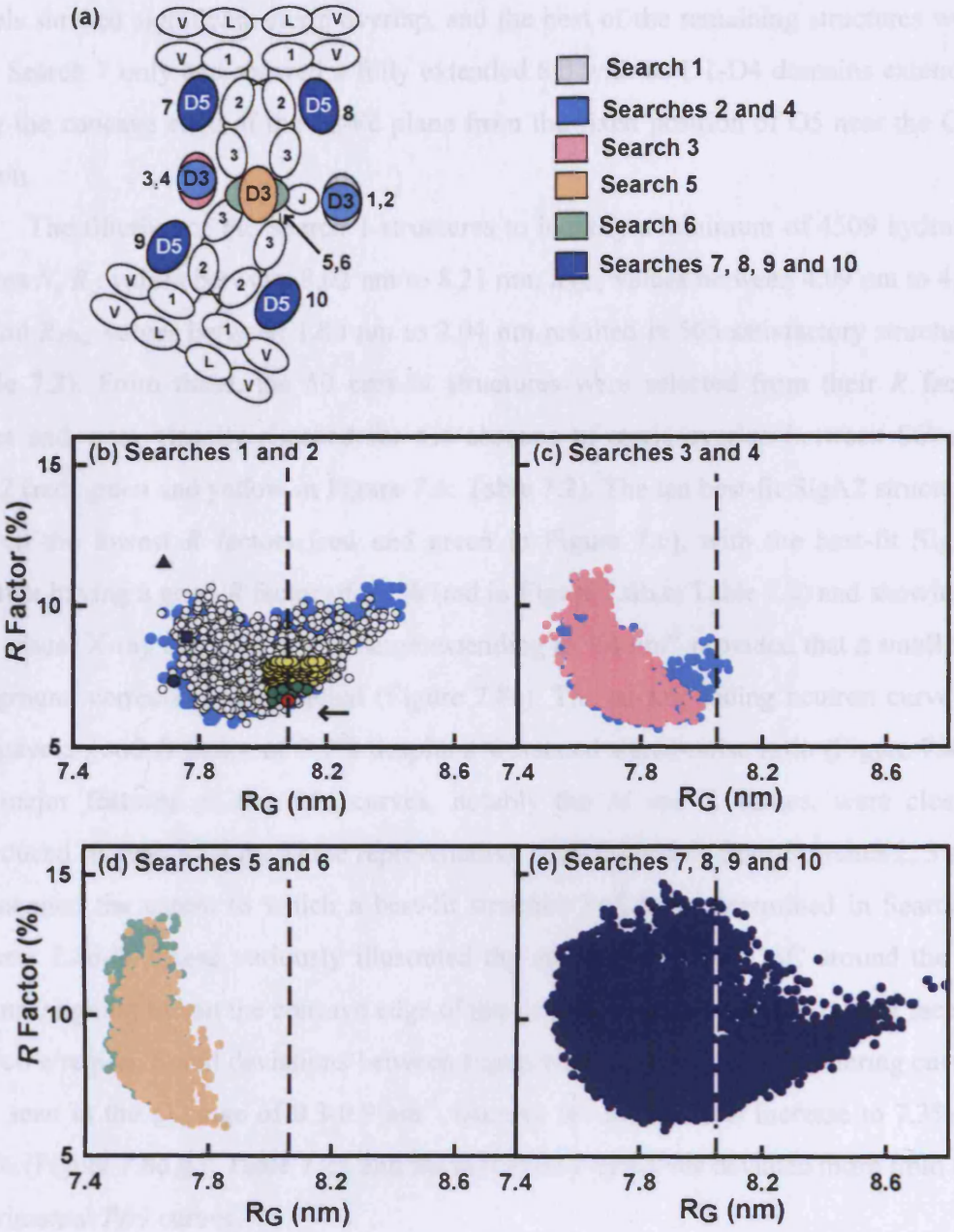

Figure 7.7 Analyses of all ten searches from the first-cycle of the SIgA2 modelling

(a) The fixed location of the SC D3 or D5 relative to dIgA1 is shown for each search labelled 1 to 10 (as in Figure 7.6a). (b-e) Comparison of the R_G and R factor values for the 1500 models from Search 1 (open circles, as in Figure 7.6b), Search 2 (blue), Searches 3 (pink), 4 (blue), 5 (orange), 6 (green), and 7-10 (dark blue). The axes remain constant for comparison. The dashed lines correspond to the experimental R_G value.

models showed significant steric overlap, and the best of the remaining structures were from Search 7 only and showed a fully extended SC with its D1-D4 domains extended along the concave edge of the Fc-Fc plane from the fixed position of D5 near the C_H2 domain.

The filtering of the Search 1 structures to identify a minimum of 4509 hydrated spheres N, R_G values between 8.02 nm to 8.21 nm, $R_{XS\text{-}1}$ values between 4.09 nm to 4.43 nm and $R_{XS\text{-}2}$ values between 1.80 nm to 2.04 nm resulted in 505 satisfactory structures (Table 7.2). From these, the 50 best-fit structures were selected from their R factor values and were visually checked for the absence of steric overlap between SC and dIgA2 (red, green and yellow in Figure 7.6; Table 7.2). The ten best-fit SIgA2 structures showed the lowest R factors (red and green in Figure 7.6), with the best-fit SIgA2 structure having a good R factor of 5.1% (red in Figure 7.6b,c; Table 7.2) and showing a good visual X-ray curve fit in a Q range extending to 1.4 nm^{-1} provided that a small flat background correction was applied (Figure 7.8a). The corresponding neutron curve fit also gave a good R factor of 7.4% despite a worsened signal-noise ratio (Figure 7.8b). The major features of the $P(r)$ curves, notably the M and L values, were closely reproduced (Figure 7.8 a,b). Three representative poor-fit models from Searches 2, 3 and 5 illustrated the extent to which a best-fit structure had been determined in Search 1 (Figures 7.8d-l). These variously illustrated the effect of wrapping SC around the Fc regions, aligning SC on the concave edge of the Fc regions, or placing SC on one face of the Fc-Fc region. Small deviations between experimental and modelled scattering curves were seen in the Q range of 0.3-0.9 nm^{-1}, causing the R factors to increase to 7.3% to 11.5% (Figure 7.8d,g,j; Table 7.2), and the modelled $P(r)$ curves deviated more from the experimental $P(r)$ curves.

A second cycle of modelling was performed to investigate whether the randomised positioning of the four Fab regions in SIgA2 would improve the Search 1 fit. Here, the best-fit location for SC against the Fc-Fc region from Search 1 (Figure 7.8c) was held fixed while the Fab regions were varied. Constrained modelling was performed to follow that previously used for monomeric IgA2m(1) (Section 7.4.3; Furtado *et al.*, 2004). A total of 3000 SIgA2 structures were generated, which were shown to test a

Filter for best-fit models	Models	Hydrated spheres (N)	X-ray R_G (nm)	X-ray R_{XS-1} (nm)	X-ray R_{XS-2} (nm)	Neutron R_G (nm)	X-ray R factor (%)	Neutron R factor (%)	$s^{\circ}_{20,w}$ (S)
Search 1 fits									
None	1500	4395-4737	7.68-8.42	3.70-4.72	1.46-2.04	6.64-7.25	5.0-10.6	6.5-12.0	n.a.
N, R_G, R_{XS-1}, R factor	50	4576-4709	8.02-8.20	4.16-4.42	1.80-1.98	6.93-7.08	5.1-7.7	6.9-8.9	n.a.
N, R_G, R_{XS-1}, R factor	10	4641-4705	8.03-8.14	4.25-4.37	1.81-1.92	6.93-7.05	5.1-7.1	7.4-8.2	11.3-11.9
Best-fit (Figure 6.8c)	1	4688	8.08	4.25	1.91	6.93	5.1	7.4	11.9
Poor-fit (Figure 6.8f) ●	1	4646	7.72	3.89	1.79	6.72	7.3	10.2	11.6
Poor-fit (Figure 6.8i) ■	1	4654	7.76	3.74	1.92	6.68	8.9	11.3	12.2
Poor-fit (Figure 6.8l) ▲	1	4593	7.69	3.53	1.97	6.56	11.5	13.3	12.2
Randomised Fab fits									
None	3000	2911-4641	4.65-8.76	0.05-4.42	0.55-2.59	4.30-7.63	6.7-33.8	8.0-33.5	n.a.
N, R_G, R_{XS-1}, R factor	31	4552-4640	7.95-8.30	4.00-4.37	1.48-1.84	7.09-7.34	7.3-9.3	8.1-10.1	n.a.
Experimental values	n.a.	4757	8.13 ± 0.10	4.22 ± 0.09	1.93 ± 0.03	7.57	n.a.	n.a.	12.1

Table 7.2 Summary of the X-ray and neutron modelling fits for the SIgA2 solution structure

n.a. – not available

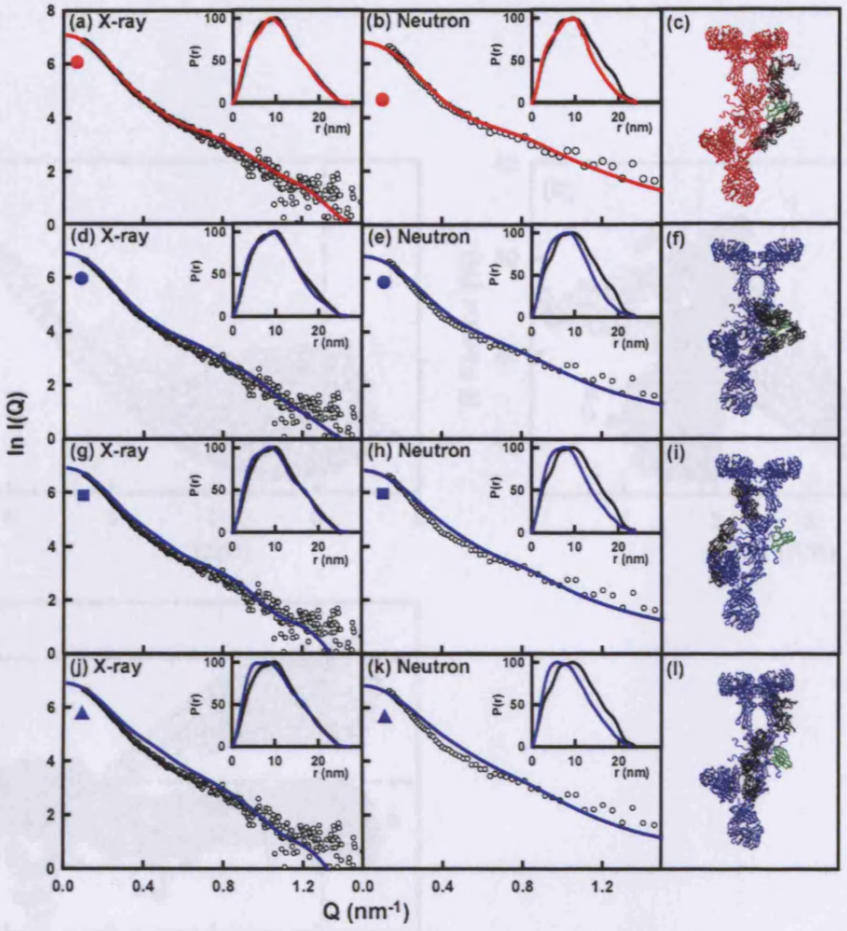

Figure 7.8 X-ray and neutron fits for the best-fit SIgA2 model and three representative poor-fit SIgA2 models.

(a,b) The X-ray and neutron experimental *I(Q)* and *P(r)* curves are in black, and the modelled best-fit SIgA2 curves are in red. The experimental *I(Q)* curve is shown in the main panel while the *P(r)* curve is shown as an inset in the top right corner. (c) The best-fit model from Search 1 is shown as a ribbon trace with the two IgA2 monomers shown in red, the J chain in green and the five domains of SC in black. (d-l) The corresponding X-ray and neutron curve fits are shown for three poor-fit SIgA2 models: (d,e,f) from Search 2 where SC is wrapped about the centre of the Fc-Fc region at the J chain between the two IgA2 monomers, (g,h,i) from Search 3, where SC is on the other (concave) edge of the Fc-Fc dimer compared to the best-fit SIgA2 model and (j,k,l) from Search 5, where SC is positioned on the central face of the Fc-Fc dimer. Other details follow those in (a,b,c). The colour and shape coding follow that in Figure 7.6.

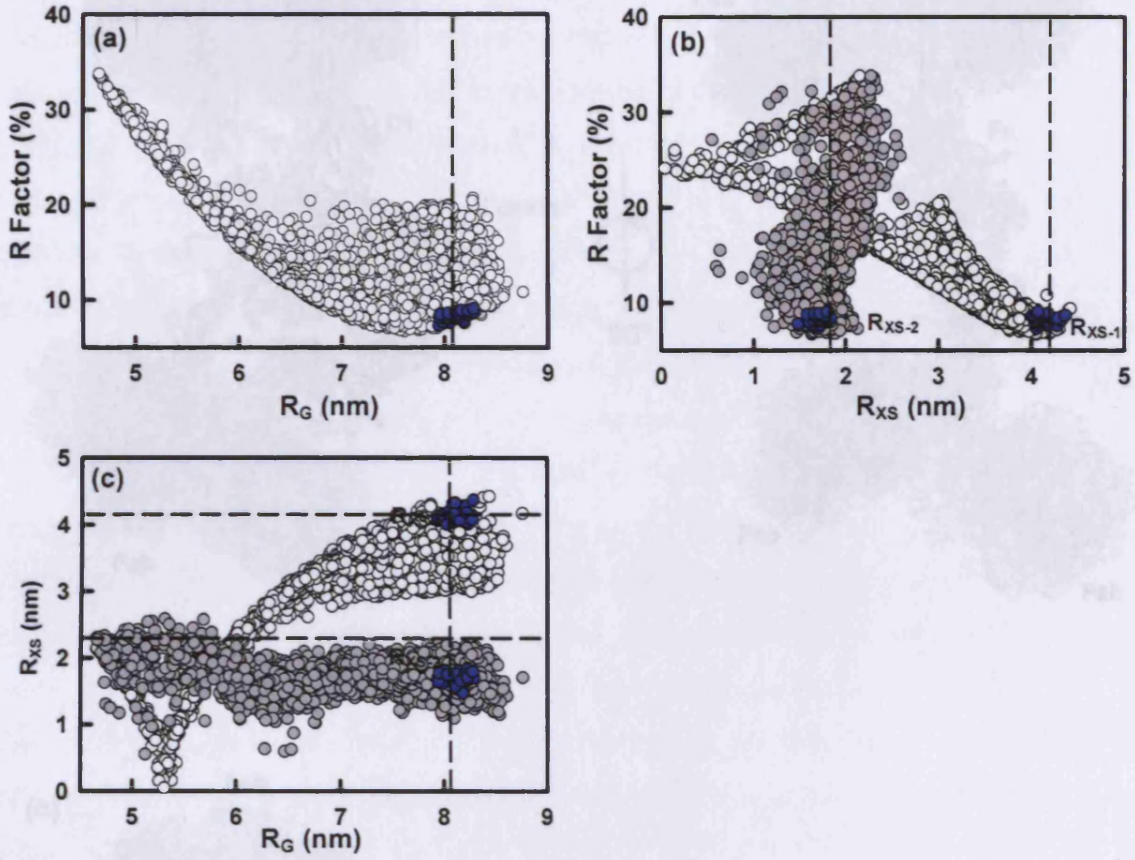

Figure 7.9 Analysis of the SIgA2 Fab reorientation search from the second-cycle of modelling fits

Comparison of the X-ray *R* factors with the X-ray R_G values for 3000 SIgA2 models. The parameters of the 31 best-fit SIgA2 models are shown in blue. The vertical dashed lines indicate the experimental X-ray R_G, R_{XS-1} and R_{XS-2} values.

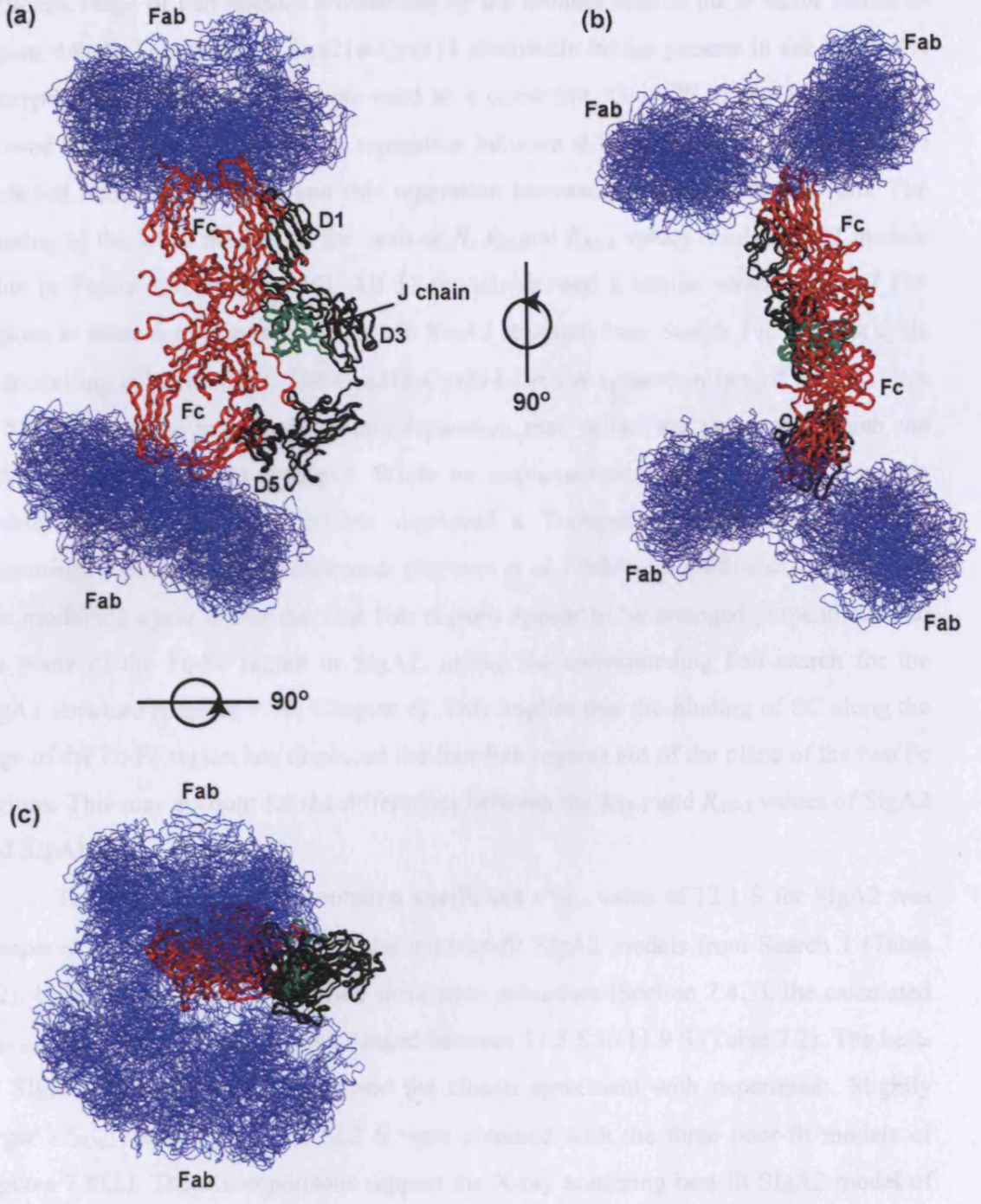

Figure 7.10 Superimposition of the 31 best-fit models for SIgA2 from the Fab reorientation search from the second-cycle of modelling fits

The SIgA2 structures are shown in three views, in which the Fab regions are shown in blue, the two Fc regions are shown as red ribbons, J chain is shown in green and SC is shown in black with the positions of D1, D3 and D5 indicated as shown.

sufficient range of Fab random orientations by the minima seen in the R factor values of Figure 7.9a,b. The inter-Fab Cys214-Cys214 disulphide bridge present in the IgA2m(1) allotype (Figure 7.1) was no longer used as a constraint. Only 79 of the 3,000 models showed a Cys214-Cys214 Cα-Cα separation between 0.14 nm to 0.65 nm (where the preferred value is 0.56 nm), and this separation increased maximally to 11.5 nm. The filtering of the 3,000 models on the basis of N, R_G and $R_{XS\text{-}1}$ values resulted in 31 models (blue in Figure 7.9; Figure 7.10). All 31 models showed a similar arrangement of Fab regions to those determined in the best-fit SIgA2 structure from Search 1 in the first cycle of modelling (Figures 7.8c). The Cys214-Cys214 Cα-Cα separation ranged from 2.5 nm to 5.6 nm, and the magnitude of this separation may reflect the presence of both the IgA2m(1) and IgA2m(2) allotypes. While no improvement in the R factors was seen (Table 7.2), all the SIgA2 models displayed a T-shaped structure similar to that determined for the IgA2m(1) monomer (Furtado *et al.*, 2004). Of particular interest from this modelling cycle is that the four Fab regions appear to be arranged perpendicular to the plane of the Fc-Fc region in SIgA2, unlike the corresponding Fab search for the SIgA1 structure (Figures 7.10; Chapter 6). This implies that the binding of SC along the edge of the Fc-Fc region has displaced the four Fab regions out of the plane of the two Fc regions. This may account for the differences between the $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ values of SIgA2 and SIgA1/dIgA1 (Table 7.1).

The experimental sedimentation coefficient $s^{\circ}_{20,w}$ value of 12.1 S for SIgA2 was compared with the $s^{\circ}_{20,w}$ values of the ten best-fit SIgA2 models from Search 1 (Table 7.2). Using a previously established simulation procedure (Section 7.4.3), the calculated $s^{o}_{20,w}$ values for the SIgA2 models ranged between 11.3 S to 11.9 S (Table 7.2). The best-fit SIgA1 model with 11.9 S showed the closest agreement with experiment. Slightly larger $s^{\circ}_{20,w}$ values of 11.6 to 12.2 S were obtained with the three poor-fit models of Figures 7.8f,i,l. These comparisons support the X-ray scattering best-fit SIgA2 model of Figure 7.8c. The superimposition of the 50 best-fit SIgA2 structures is shown in Figure 7.11, and the ten best-fit SIgA2 structures are displayed in Figure 7.12.

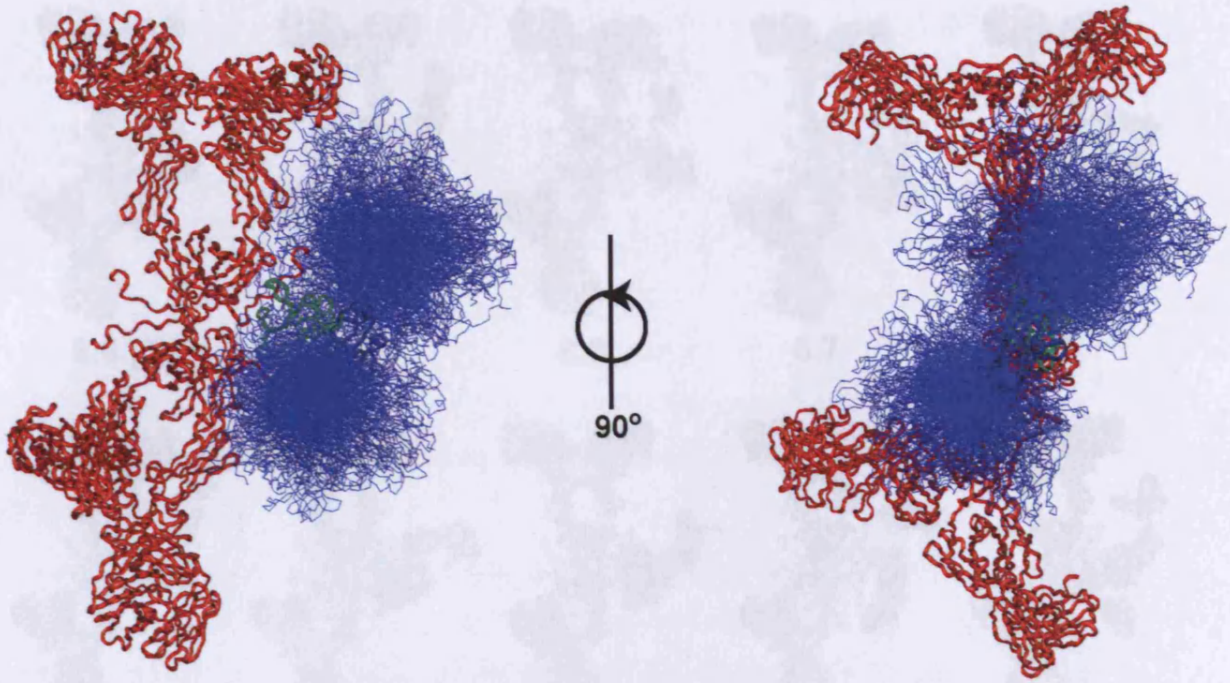

Figure 7.11 Superimposition of the 50 best-fit SIgA2 models from Search 1

The 50 best-fit SIgA2 models (● in Figure 7.6) are shown superimposed in two orthogonal views that are face-on to (left) and in the plane of the Fc regions (right). dIgA2 is shown in red, J chain is shown in green and the α-carbon traces of SC are shown in blue, except for the fixed domain D3 which is shown as a ribbon.

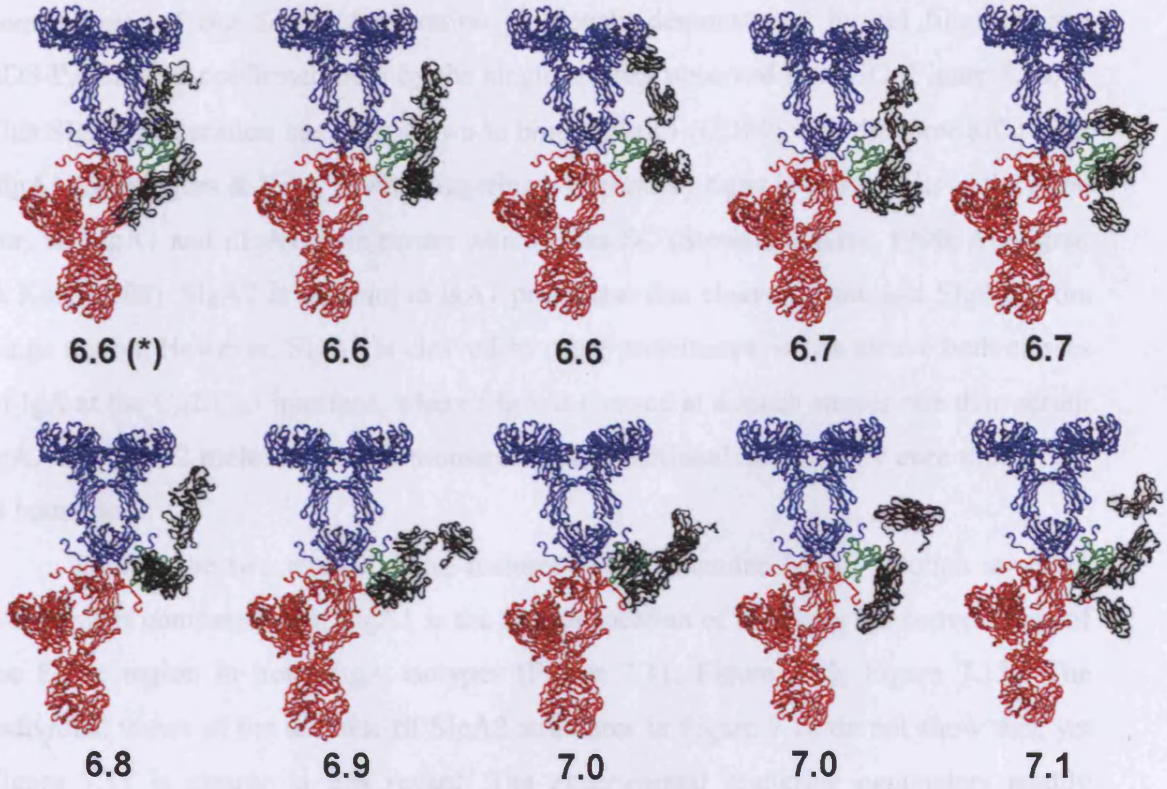

Figure 7.12 The ten best-fit SIgA2 models from Search 1

The ten best-fit models are shown (● in Figure 7.6) with the two IgA2 monomers in red and blue, each pair being shown in the same orientation. The variable location of SC in the models is shown in black. The J chain is shown in green. The first model (*) corresponds to the best-fit SIgA2 model (● in Figure 7.6; Figure 7.8c). The *R* factor of each model is shown beneath each one.

7.3 Discussion

Given that SIgA is one of the most important components of mucosal immunity, the solution structure for SIgA2 at medium resolution shows not only how this is assembled from its components, but also completes our structural understanding of the major forms of human mucosal IgA by its comparison with SIgA1. As for SIgA1, the homogeneity of our SIgA2 preparation previously demonstrated by gel filtration and SDS-PAGE was confirmed here by the single species observed by AUC (Figure 7.5b,c). This SIgA2 preparation has been shown to bind to FαR1 (CD89) with the same affinity as SIgA1 (Mazengera & Kerr, 1990), triggering a respiratory burst in neutrophils in the same way as SIgA1 and dIgA1 from serum which lacks SC (Stewart & Kerr, 1990; Almogren & Kerr, 2008). SIgA2 is resistant to IgA1 proteinase that cleaves serum and SIgA1 at the hinge region. However, SIgA2 is cleaved by other proteinases, which cleave both classes of IgA at the C_H2/C_H3 interface, where SIgA is cleaved at a much slower rate than serum IgA. The SIgA2 molecule thus demonstrated full functional accessibility even though SC is bound to it.

One of the two most striking features in the extended SIgA2 solution structure when this is compared with SIgA1 is the similar location of SC along the convex edge of the Fc-Fc region in both SIgA isotypes (Figure 7.11; Figure 7.12; Figure 7.13). The individual views of the ten best-fit SIgA2 structures in Figure 7.12 do not show this, yet Figure 7.11 is clearer in this regard. The experimental scattering parameters readily showed the similarity of their extended solution structures from the R_G/R_O ratios and the M and L values in the $P(r)$ curves (Table 7.2), with small differences being attributable to the four shorter hinges in SIgA2 (Figure 7.2). Even though no data were available for dIgA2 to act as a bench mark, good scattering modelling fits were nonetheless obtained by starting from the dIgA1 solution structure with the two monomers orientated end-to-end through a slightly bent arrangement of their Fc regions (Figures 7.8a,b; Figure 7.13). As for SIgA1, the curve fits again located the SC structure along the outermost convex edge of the Fc-Fc region, although it is emphasized that the specific interactions of the five SC domains with the Fc-Fc region could not be identified at the resolution of this structural determination method. As for SIgA1, the best-fit location for SC within SIgA2 was identified through the comparison of the modelled R_G values with the experimental

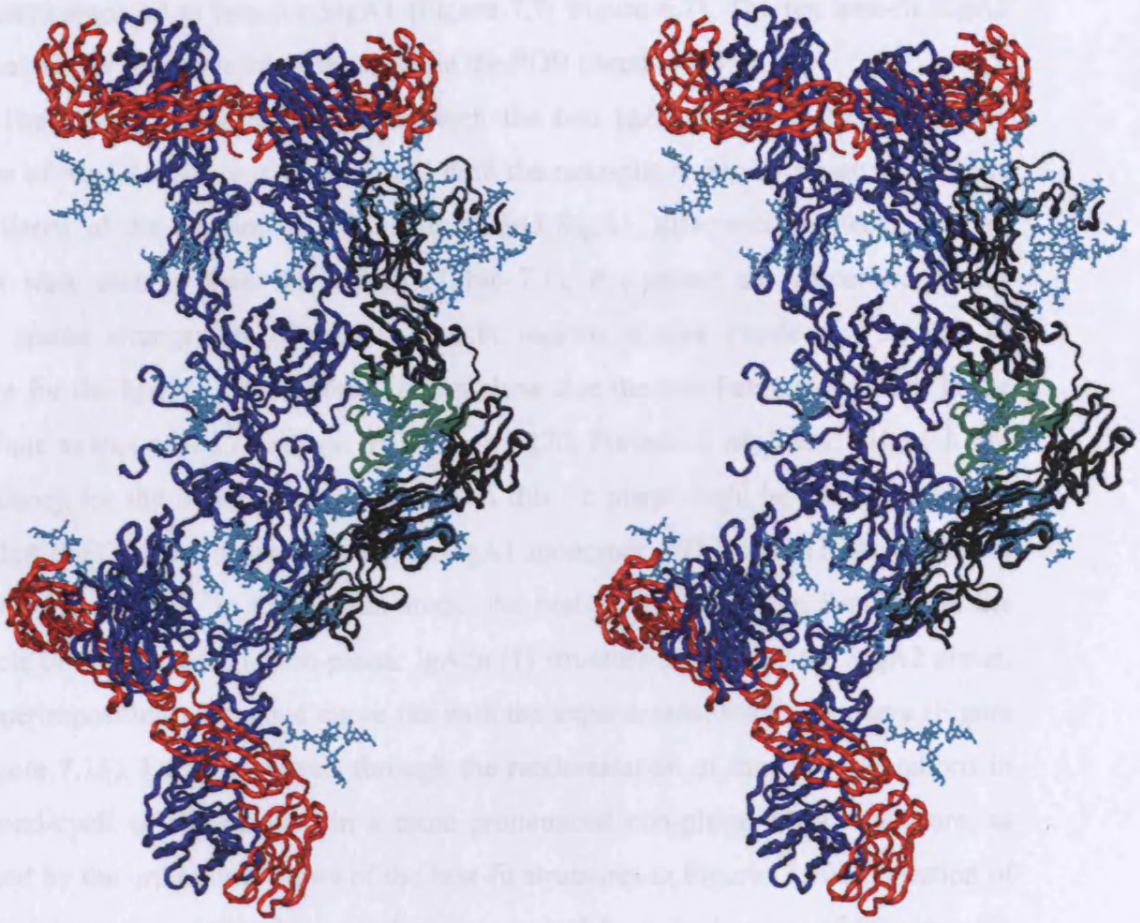

Figure 7.13 Stereo view of the solution structure of SIgA2

The best-fit SIgA2 model (Figure 7.8c; Figure 7.12) is shown as a stereo pair. The heavy chains (blue), light chains (red), J chain (green) and SC (black) are shown. N-linked oligosaccharides are shown in cyan.

value of 8.13 nm (Figure 7.6; Figure 7.7). Many earlier hypothetical models for SIgA that placed SC on one face of the Fc-Fc region (Heremans, 1974; Kerr, 1990) resulted in R_G values that were too low when modelled, and this was more clearly demonstrated by the SIgA2 fits compared to that for SIgA1 (Figure 7.7; Figure 6.7). The ten best-fit SIgA2 models of Figure 8(a) have been deposited in the PDB (Section 7.4.4).

The other striking difference between the two IgA isotypes is the non-planar structure of SIgA2 when this is compared with the near-planar SIgA1 structure. Despite the similarity of the location of SC in SIgA2 and SIgA1, differences between the two isotypes were seen in their R_{XS} values (Table 7.1). R_{XS} values are influenced by the relative spatial arrangement of the Fab and Fc regions in Abs. Previously, the best-fit structure for the IgA2m(1) monomer did not show that the two Fab regions were in the same plane as that of the Fc region (PDB code 1R70; Furtado *et al.*, 2004; Figure 1.7b). No tendency for the two Fab regions to lie in this Fc plane could be discerned in that earlier IgA2m(1) study, unlike that for the IgA1 monomer (PDB code 1IGA; Boehm *et al.*, 1999; Figure 1.7a). In the present study, the best-fit structure from Search 1 in the first-cycle of fits carried this non-planar IgA2m(1) structure across into the SIgA2 dimer. This superimposition gave good curve fits with the experimental scattering curve (Figure 7.8; Figure 7.13). Further analysis through the randomisation of the Fab orientations in the second-cycle of fits resulted in a more pronounced non-planar SIgA2 structure, as evidenced by the orthogonal views of the best-fit structures in Figures 7.10. Inspection of these models suggested that this non-planarity resulted from the location of SC along the convex edge of the Fc-Fc region, which leads to crowding between the D1 and D5 domains of SC and the adjacent two Fab regions, caused by the shorter hinge region in IgA2. In turn, this effect appears to cause the four Fab regions to be displaced out of the plane of the Fc-Fc region.

A non-planar SIgA2 structure with SC located at one edge of its Fc-Fc region might result in a different mode of interaction with cell surface receptors compared to SIgA1, most notably to FcαRI which is responsible for IgA-mediated phagocytosis, oxidative burst and Ab-dependent cellular cytoxicity amongst other roles (Monteiro & van de Winkle, 2003). SIgA2 consistently triggers a slightly slower (though similarly sized) neutrophil respiratory burst than SIgA1 (Mazengera & Kerr, 1990). Little is known

about the differences in effector functions of the two SIgA isotypes. Like SIgA1, the SIgA2 structure shows that receptor binding would occur independently of Ag binding as the Fc and Fab regions are well separated. Visualisation of the best-fit SIgA2 structure with the FcαRI crystal structure (PDB code 1OW0; Herr *et al.*, 2003b) shows that two FcαRI molecules will potentially bind at the junction between the C_H2 and C_H3 domains per Fc region, making this four per IgA2 dimer as Sc blocks two sites. This SIgA2 interaction with FcαRI is similar to that seen for SIgA1 (Figure 6.13). It has been deduced that a 1:1 stoichiometry of SIgA binding to FcαRI on cell surfaces will be the same for both SIgA isotypes, despite the difference in the planarity of their structures.

The SIgA2 solution structure provides insight into the way that SIgA2 is formed from dIgA2 and SC. The J chain is presumed to be located at the convex bend between the two Fc regions, even though the J chain could not be located at the resolution of scattering modelling (Figure 7.13; Chapter 5). This is consistent with the immunologically hidden location of the J chain (Heremans, 1974). Unbound SC shows a J-shaped domain arrangement which will have to unfold into an extended domain arrangement in order to bind to the IgA dimer. As for SIgA1, the SIgA2 structure supports a 'zipper effect' binding mechanism of SC to dimeric IgA (Crottet & Corthésy, 1999). A non-planar arrangement of the Fab regions relative to the Fc-Fc regions in the IgA2 dimer makes the Fc-Fc region accessible to the pIgR. The pIgR is thought to still be able to interact with the IgA2 dimer prior to transcytosis. However, exact dIgA2 residues involved in the pIgR interaction are currently unknown.

The location of the SC D1-D5 domains along the edge of the Fc-Fc regions in SIgA1 provides SIgA1 with a greater ability to resist proteolysis in the harsh environment of mucosal surfaces. SC on its own is readily degraded to half its size (Heremans, 1974; Beale, 1985), but this does not occur when SC is bound to SIgA1. Proteases cleave monomeric IgA1 and dIgA1 more readily than SIgA1 (Lindh, 1975; Almogren *et al.*, 2003). However, the SC in SIgA2 is cleaved at a similar rate to free SC, which is much more rapidly than SC in SIgA1 (Almogren *et al.*, 2006). The SIgA2 model explains this decrease in resistance (when compared to SIgA1) in terms of the greater exposure of SC along the convex edge of the Fc-Fc region. For reason of its non-planarity, SIgA2 is no longer shielded by the Fab regions in the manner described for SIgA1 (Section 6.3). Free

SC is susceptible to cleavage between D3 and D4 for reason of the long peptide linker between these (Chapter 4). The non-covalent link between SC and the C_H2 domain in SIgA2 (Almogren *et al.*, 2007) is expected to contribute to this susceptibility. It is possible that the notably heavier glycosylation of IgA2 at 16 glycosylation sites at Asn166, Asn263, Asn337 and Asn459 on the two Fc regions represents an evolutionary attempt to protect SIgA2 from proteolytic cleavage (Figure 7.13). However the evolution of the SIgA1 isotype has provided a more satisfactory solution to this issue.

The non-planar solution structure for SIgA2 with SC located to the edge of the Fc-Fc region has rationalised the existence of the two human IgA isotypes. The isotypes have characteristic distributions within mucosal surfaces, with slightly more SIgA1 than SIgA2 except in colonic mucosa. SIgA2 is most similar to IgA homologues in other species, where SIgA1 is only seen in higher apes and humans. The main difference between the SIgA isoforms was previously believed to be in the hinge region. The present work indicates that the solution structures of SIgA2 and SIgA1 possess non-planar and planar structures respectively. These differences mean that, in the harsh environment of mucosal secretions, SIgA1 and SIgA2 are able to bind to different Ags and have varied susceptibilities to mucosal proteases, therefore increasing the efficiency of the mucosal immune system. The outcome of this study has clarified the further work needed to understand SIgA assembly and function, in particular the molecular details of the binding motifs within IgA2 and SC within SIgA2.

7.4 Materials and Methods

7.4.1 Preparation and composition of SIgA2

Native SIgA2 (expected to contain λ and κ class light chains and both the SIgA2m(1) and SIgA2m(2) allotypes) was isolated from human colostrum provided by consenting healthy volunteers and purified using three-step procedure of thiophilic chromatography, jacalin-agarose lectin affinity chromatography and gel filtration (Sun *et al.*, 2005; Almogren & Kerr, 2007). SIgA2 preparations were provided by Dr Adel Almogren and Professor Michael A. Kerr. Briefly, the thiophilic resin extracted the immunoglobulins (SIgA1, SIgA2, secretory IgM and free SC), and the jacalin-agarose column specifically bound to SIgA1 to leave SIgA2, polymeric SIgA2, secretory IgM, and free SC in the eluate. A Sephyacryl S-300 gel filtration column separated these as three well-resolved peaks containing secretory IgM and polymeric SIgA2, dimeric SIgA2 and free SC. The dimeric SIgA2 fractions were collected and concentrated. The purity of the dimeric SIgA2 sample was ensured by FPLC (where a single symmetrical SIgA2 peak was produced), SDS-PAGE and Western blots using anti-human α chain, anti-human κ and λ light chain and anti-human SC antibodies. Samples were also checked by reducing and non-reducing SDS-PAGE to confirm sample integrities before and after data collection. For X-ray scattering and AUC experiments, SIgA2 preparations were studied in Dulbecco's phosphate buffer saline (PBS; 12.5 mM sodium phosphate, 140 mM NaCl, pH 7.4) supplemented with EDTA and sodium azide as anti-bacterial preservatives (0.5 mM EDTA, 0.02% NaN_3). For neutron scattering, the buffer was Dulbecco's PBS as above, and made up in 100% % ^2H_2O. The neutron samples were dialysed at 6°C into this 100% ^2H_2O buffer for 36 h with four buffer changes.

For the data analyses, the composition of SIgA2 was taken to correspond to two IgA2m(1) monomers, J chain and SC. The α-chain sequence for the C_H1, C_H2 and C_H3 domains for IgA2m(1) was taken from SWISSPROT (database code: P01877), while the V_H, V_L and C_L sequences were taken as stated in the previous study of monomeric IgA2m(1) (Furtado *et al.*, 2004). The N-linked oligosaccharides at Asn169, Asn253, Asn327 and Asn449 were represented by biantennary complex-type carbohydrates (Figure 7.1; Figure 7.2). The J chain and SC compositions were taken from that used previously (Section 4.4.1; Section 5.4.1). Based on this, SIgA2 with four κ light chains has a

molecular mass of 424 kDa, an unhydrated volume of 537 nm^3, a hydrated volume of 711 nm^3 (based on a hydration of 0.3g H_2O/g of glycoprotein and an electrostricted volume of 0.0245 nm^3 per bound water molecule), an absorption coefficient (1%, 1 cm path length) of 12.2, and a $\bar{v}$ of 0.721 ml/g (Perkins, 1986). The corresponding mass for SIgA2 with four λ light chains is 417 kDa.

7.4.2 X-ray and neutron scattering and ultracentrifugation data

X-ray scattering data were obtained in two data sessions by Dr. Patricia Furtado at the European Synchrotron Radiation Facility (ESRF), Grenoble, France (Narayanan *et al.*, 2001; Section 2.2.3.2) on the Beamline ID02, both in single-bunch mode to reduce the incident flux. Storage ring currents ranged from 11 mA to 18 mA with a ring energy of 6.0 GeV. The sample to detector distance of 3.0 m yielded a Q range of 0.09 to 2.0 nm^{-1}. Samples at 0.32 mg/ml to 1.94 mg/ml were contained in water-cooled Perspex cells at 15°C, of path thickness 1 mm and mica windows of thickness 25 μm. Samples were measured in 10 time frames, each of which lasted for 0.1 sec, 1 sec or 2 sec, in order to assess and eliminate possible radiation damage effects. Neutron scattering data for SIgA2 were obtained by Dr. Patricia Furtado on Instrument LOQ at the pulsed neutron source ISIS at the Rutherford Appleton Laboratory, Didcot, U.K. (Heenan *et al.*, 1997; Section 2.2.3.3). The proton beam current was approximately 180 μA and data acquisition times were between 4 to 12 h at a concentration of 1.00 mg/ml, measured at 15°C over a Q range of 0.06 to 2.3 nm^{-1}. Other details including the data reduction procedure, the Guinier analyses and the calculation of the distance distribution function $P(r)$ using GNOM are as described previously (Section 2.2.4). For GNOM, the SIgA2 X-ray $I(Q)$ curve contained 488 data points between Q values of 0.09 to 2.00 nm^{-1} and was fitted with D_{max} set as 25 nm. The SIgA2 neutron $I(Q)$ curve contained 71 data points between Q values of 0.13 to 2.3 nm^{-1} and was fitted with D_{max} set as 24 nm.

AUC analyses for SIgA2 were performed at 20°C on a Beckman XL-I instrument equipped with an AnTi50 rotor, recording both interference and 280 nm absorbance scans. SE data sets were recorded for SIgA2 concentrations of 0.90 mg/ml, 0.60 mg/ml and 0.32 mg/ml over 45 h. Data were acquired at rotor speeds of 7,000 r.p.m., 9,000 r.p.m., 11,000 r.p.m. and 14,000 r.p.m. until equilibrium had been reached at each speed

as shown by the perfect overlay of runs measured at 5 h intervals. A final run at 42,000 r.p.m. was performed at the end in order to determine the background levels for use in curve fits. Molecular mass values were determined on the assumption of a single species using ORIGIN v4.1 (Microcal). The buffer density was calculated to be 1.00543 g/ml from its composition using SEDNTERP (Laue *et al.*, 1992), and the viscosity was taken as 0.01002 cp. SV data were acquired at SIgA2 concentrations of 0.93 mg/ml, 0.57 mg/ml and 0.40 mg/ml. Data were collected over 16 h at rotor speeds of 10,000 r.p.m., 15,000 r.p.m., 20,000 r.p.m., 25,000 r.p.m. and 30,000 r.p.m. Data were analysed using DCDT+ $g(s^*)$ time-derivative analyses (Philo, 2006) and SEDFIT v9.4 (Schuck, 2000) based on the continuous *c(s)* distribution model with a resolution set as 150, while the cell meniscus and bottom, the frictional ratio of 1.53, and the partial specific volume $\bar{v}$, buffer density and viscosity were held fixed, and the baseline was allowed to float. Other details are as described previously (Section 2.3; Section 4.4.3).

7.4.3 Constrained modelling of SIgA2

The scattering and sedimentation data were subjected to constrained modelling in order to determine the SIgA2 solution structure (Perkins *et al.*, 2008). One constraint was the solution structure of dIgA1, which was determined by constrained modelling starting from the monomer of IgA1 (PDB codes 1IGA and 2QTJ; Boehm *et al.*, 1999; Chapter 5; Bonner *et al.*, 2008). The Fc-Fc regions in this dimer was used as the template for superimposition of the Fc region in the solution structure for monomeric IgA2m(1) (PDB code 1R70; Furtado *et al.*, 2004) to create a starting model for dIgA2. The sequence numbering follows that of IgA1 Bur for ease of comparison (Boehm *et al.*, 1999; Figure 7.2). N-linked oligosaccharides were added to Asn166, Asn253, Asn327 and Asn449 as bianntennary complex-type with a $Man_3GlcNAc_2$ core and two NeuNAc.Gal.GlcNAc antennae. The other constraint was the use of 5,000 conformationally randomised full homology models for SC (including seven oligosaccharide chains) that were generated by the unrestrained linker search modelling for SC (Chapter 4; Bonner *et al.*, 2007). The J chain in the dIgA1 model was arbitrarily represented by domain D4 of SC and was located between the two Fc regions (Figure 5.13). Starting from this dIgA2 structure, the first cycle of searches was performed by fixing the D3 or D5 domain of the 5,000 SC

models in ten distinct locations adjacent to dIgA2, then evaluating the resulting SIgA2 models. Searches 1 and 2 were performed with D3 next to the J chain domain on the convex edge of the Fc region in which D3 was aligned with the plane of the Fc region or perpendicular to this. Searches 3 and 4 held D3 in two similar orientations on the other concave edge of the Fc region. Searches 5 and 6 held D3 in two similar orientations on top of the centre of the Fc dimeric region. Four searches 7 to 10 were performed with D5 held adjacent to the four C_H2 domains at the top of the Fc regions (Figure 7.6a; Figure 7.7a). The second search cycle randomised the hinge between the Fab and Fc regions in the best-fit SIgA2 model from the first search (Figure 7.8c). Six linker libraries with 5000 hinge conformations in each one were generated using molecular dynamics for 12-residue hinge lengths set between 2.50 nm to 3.75 nm in six 0.25 nm steps, and 3,000 of the resulting 30,000 conformations were randomly selected for the SIgA2 models. The hinge between the C_H1 and C_H2 domains is 220-CPVPPPPPCCHP-244 (Figure 7.2). Four identical linker structures were used to generate 3,000 SIgA2 models for evaluation.

Each SIgA2 coordinate model was converted into Debye spheres in order to calculate X-ray and neutron scattering curves for comparison with the experimental curves (Section 3.3.3). A cube side length of 0.531 nm in combination with a cutoff of 4 atoms consistently produced sphere models within 1% of the total unhydrated volume of 537 nm^3 calculated from its composition. The optimal total of dry spheres for SIgA2 was 3589. To add the hydration shell that is visible by X-ray scattering, the sphere models were adapted by adding spheres to the surface of the dry models using HYPRO (Ashton *et al.*, 1997) based on 0.3 g H_2O / g glycoprotein to generate a total hydrated volume of 711 nm^3. The optimal total of hydrated spheres for SIgA2 was 4747. In order to reduce the computing time to evaluate best-fit SIgA2 structures, the 5,000 SC models were first tested for steric overlap between their five domains using a 7% volume filter to leave only 1500 SC models. The ten searches thus evaluated a total of 15,000 SIgA2 models. The values of N and the modelled R_G, $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ parameters were calculated in the same Q ranges used for the experimental Guinier analyses. Models that fit within 5% of N and the X-ray Guinier values were ranked using the goodness-of-fit R factor of the model. A flat background correction of 0.2% of $I(0)$ was subtracted from the modelled X-ray scattering curve shown in Figure 7.8 (where $I(0)$ is 1000) as an empirical correction of

differences in the sample and buffer transmissions that were observed at larger Q values. No flat-background correction was required in the neutron fits. The sedimentation coefficients an $s^{\circ}_{20,w}$ were calculated from the final SIgA2 coordinate models using HYDROPRO v7c. To take into account the hydration shell surrounding SIgA2, the recommended value of 0.31 nm for the atomic element radius for all atoms was used as an empirical approximation of this.

7.4.4 SIgA2 Protein Data Bank accession code

The ten best-fit SIgA2 α-carbon co-ordinate models from Search 1 (Figure 7.12; Table 7.2) were deposited in the Protein Data Bank with the accession code 3CM9.

Chapter Eight

Conclusions

8.1 Conclusions

8.1.1 History of structural models for secretory IgA

The structure of SIgA has been under much debate over the years and many models have been proposed that differ greatly from one another. With no high-resolution structures of intact SIgA available, low-resolution structural-determination methods were employed to determine the SIgA macromolecular structure in this thesis. SIgA was first visualised by EM in 1970, where SIgA was seen to have a structure similar to an IgA monomer, yet with greater electron density in the Fab regions (Figure 1.9a; Svehag & Bloth, 1970). Structural models of SIgA were consequently proposed, with the two IgA monomers congruently on top of one another, with SC hidden between the two Fc regions (Figure 8.1a,c,d; Svehag & Bloth, 1970).

In contrast, EM models for dIgA (Figure 1.9; Svehag & Bloth, 1970; Dourmashkin *et al.*, 1971, Munn *et al.*, 1971), positioned the two IgA monomers in an end-to-end arrangement (Figure 8.1b; Munn *et al.*, 1971). The SIgA structural model was then reconsidered in 1971, based on further EM experiments (Figure 1.9c; Bloth & Svehag, 1971). Revised models of SIgA were similar to that of dIgA, with the two IgA monomers arranged in an end-to-end conformation (Figure 8.1e,h-o; Heremans, 1974). Studies to assign the disulphide bonds with SIgA and dIgA were carried out from 1986, to determine the position of the covalent linkages between the components of SIgA (Figure 8.f,h,i; Garcia-Pardo *et al.*, 1986; Underdown *et al.*, 1986; Mestecky *et al.*, 1987). It was not known whether the Fc regions of the two IgA monomers overlapped in dIgA and SIgA, or whether the J chain acts as a spacer between the Fc-Fc regions (Figure 8.1h,i,k).

An early model for the domain arrangement in free SC proposed this to be a compact cyclic structure (Figure 8.1g; Beale, 1985). However, the position of bound SC was undetectable in all these early studies, as SC is too small to be visualised by EM. However, SC was hypothesised to wrap itself around the dIgA Fc-Fc region, for reason of the mutual protection of SC and dIgA from proteolysis (Figure 8.1j,m,n; Kerr, 1990, Woof & Mestecky, 2005; Arnold *et al.*, 2007). These tentative models were common in many reviews of IgA (Figure 8.1).

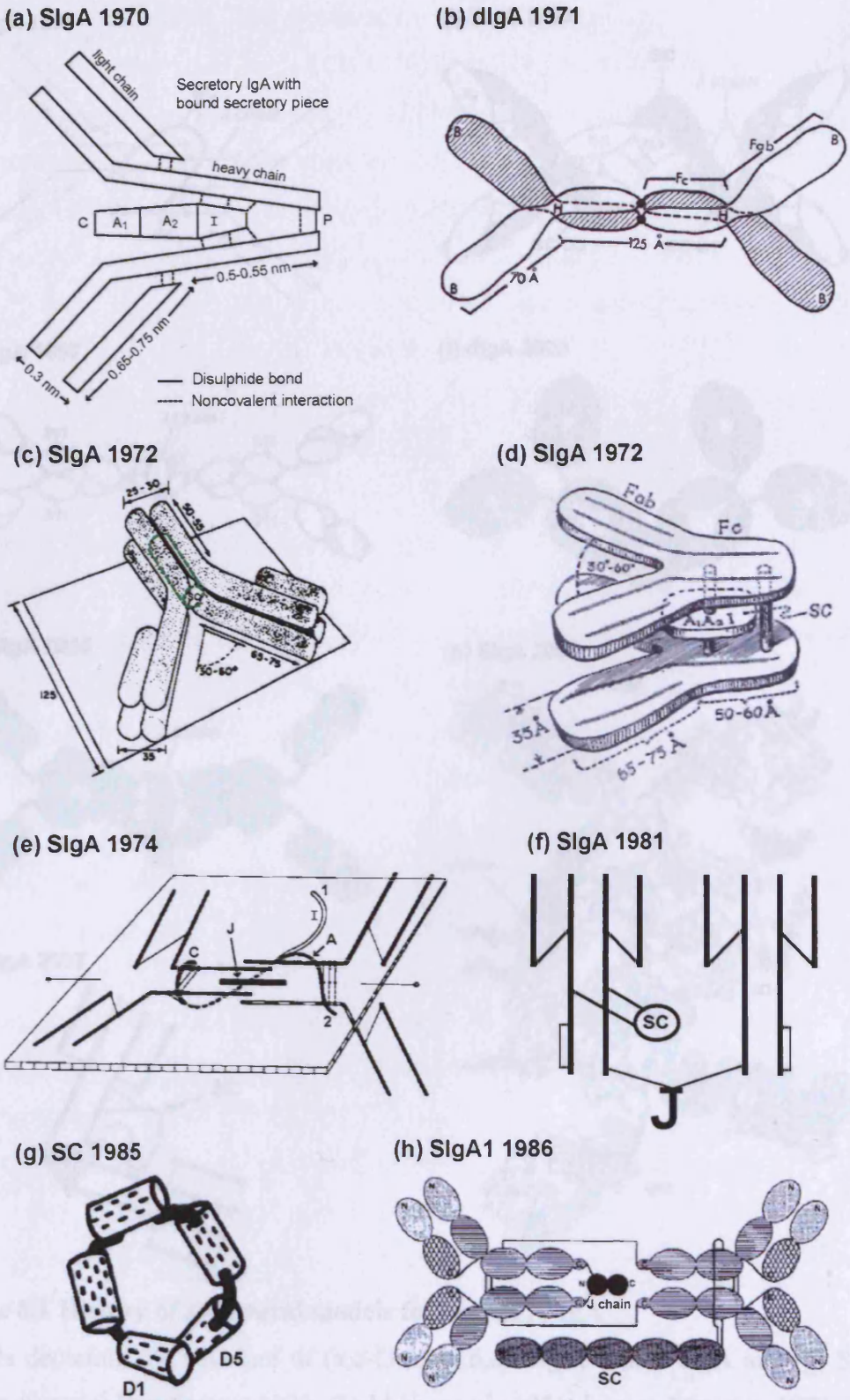

Figure 8.1 History of structural models for secretory IgA

Figure continued overleaf.

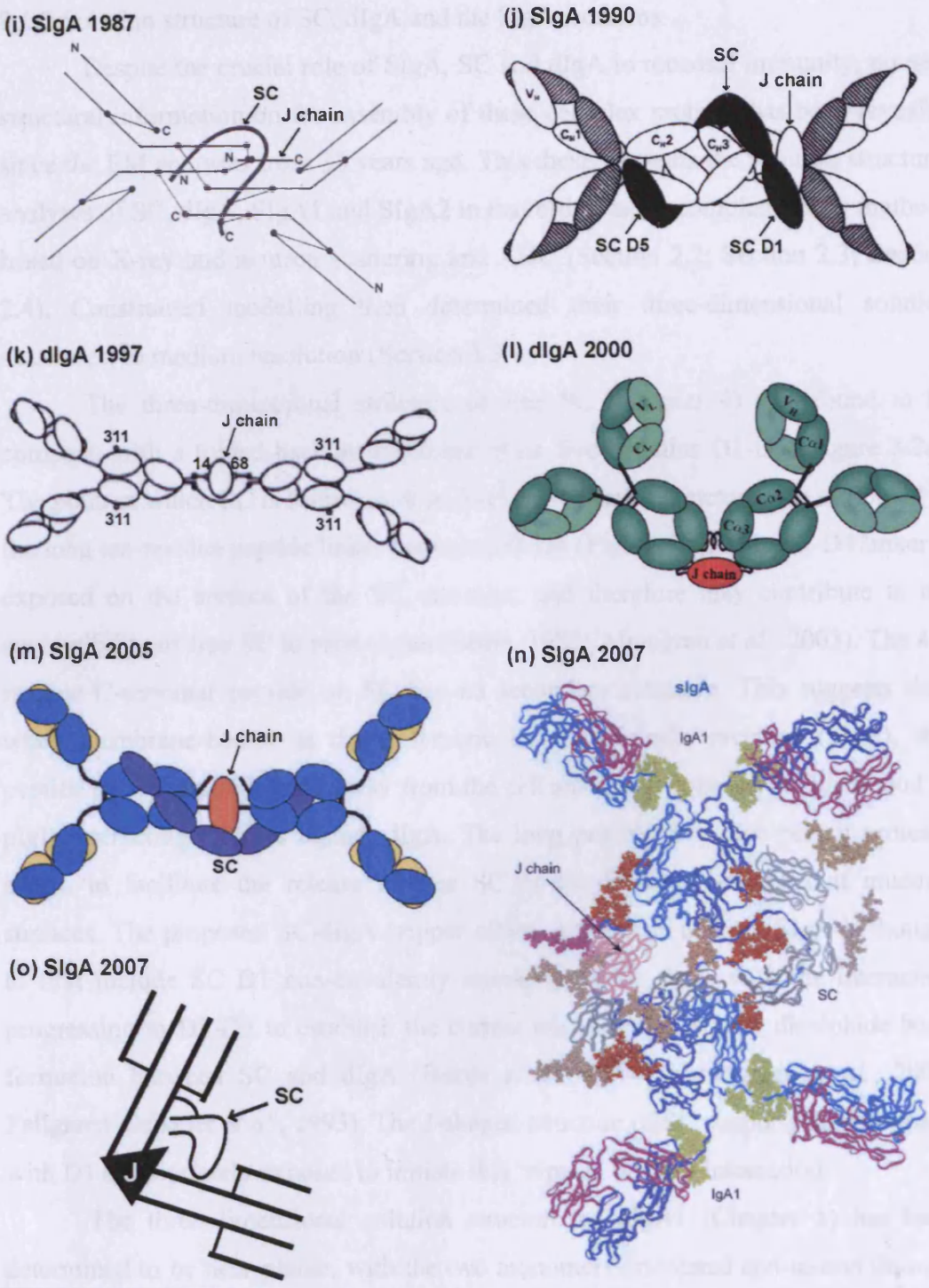

Figure 8.1 History of structural models for secretory IgA

Models depicting the structure of (a,c-f,h,i,j,m,n,o) SIgA, (b,k,l) dIgA and (g) SC. (Taken from (a) Brandtzaeg, 1970; (b) Munn *et al.*, 1971; (c) van Munster, 1972; (d) Tomasi & grey, 1972; (e) Heremans, 1974; (f) Garcia-Pardo *et al.*, 1986; (g) Beale, 1985; (h) Underdown & Schiff, 1986; (i) Mestecky *et al.*, 1987; (j) Kerr, 1990; (k) Krugman *et al.*, 1997; (l) Johansen *et al.*, 2000; (m) Woof & Mestecky, 2005; (n) Arnold *et al.*, 2007; (o) Brandztaeg, 2007).

8.1.2 Solution structure of SC, dIgA and the SIgA isoforms

Despite the crucial role of SIgA, SC and dIgA in mucosal immunity, no new structural information on the assembly of these complex proteins has been revealed since the EM analyses from 38 years ago. This thesis presents the solution structural analyses of SC, dIgA, SIgA1 and SIgA2 in that order, using complementary methods based on X-ray and neutron scattering and AUC (Section 2.2; Section 2.3; Section 2.4). Constrained modelling then determined their three-dimensional solution structures, to medium resolution (Section 3.3).

The three-dimensional structure of free SC (Chapter 4) was found to be compact, with a folded-back arrangement of its five domains D1-D5 (Figure 8.2a). The point at which SC is folded back to form the J-shaped structure was attributed to the long ten-residue peptide linker between D3-D4 (Figure 4.1). This D3-D4 linker is exposed on the surface of the SC structure, and therefore may contribute to the susceptibility of free SC to proteolysis (Beale, 1985; Almogren *et al.*, 2003). The 42-residue C-terminal peptide on SC has no secondary structure. This suggests that, when membrane-bound as the polymeric immunoglobulin receptor (pIgR), this peptide may position D1-D5 away from the cell surface to enhance the likelihood of pIgR interacting with its ligand, dIgA. The long peptide may also permit protease access to facilitate the release of free SC or the SC-dIgA complex at mucosal surfaces. The proposed SC-dIgA 'zipper effect' interaction to form SIgA is thought to first include SC D1 non-covalently interacting with dIgA, with the interaction progressing to D2-D3 to establish the correct alignment of SC for disulphide bond formation between SC and dIgA (Bakos *et al.*, 1994; Hamburger *et al.*, 2004; Fallgreen-Gebauer *et al.*, 1993). The J-shaped structure of SC supports this proposal, with D1 appropriately exposed to initiate this 'zipper' binding interaction.

The three-dimensional solution structure of dIgA1 (Chapter 5) has been determined to be near-planar, with the two monomers orientated end-to-end through their Fc regions. The two Fc regions are in close proximity to each other, and are slightly bent in their relative arrangement (Figure 8.2b), supporting the EM findings and the subsequent proposed dIgA models (Figure 1.9; Figure 8.1b,k,l). However, the J chain does not act as a spacer between the Fc regions of the two IgA monomers (Figure 8.2b; Figure 8.1k). The J chain was best positioned in dIgA1 within a crevice formed by the bent Fc-Fc region on the convex edge, although a crystal structure will be necessary to confirm this (Figure 8.2b). The central location of the J chain may be

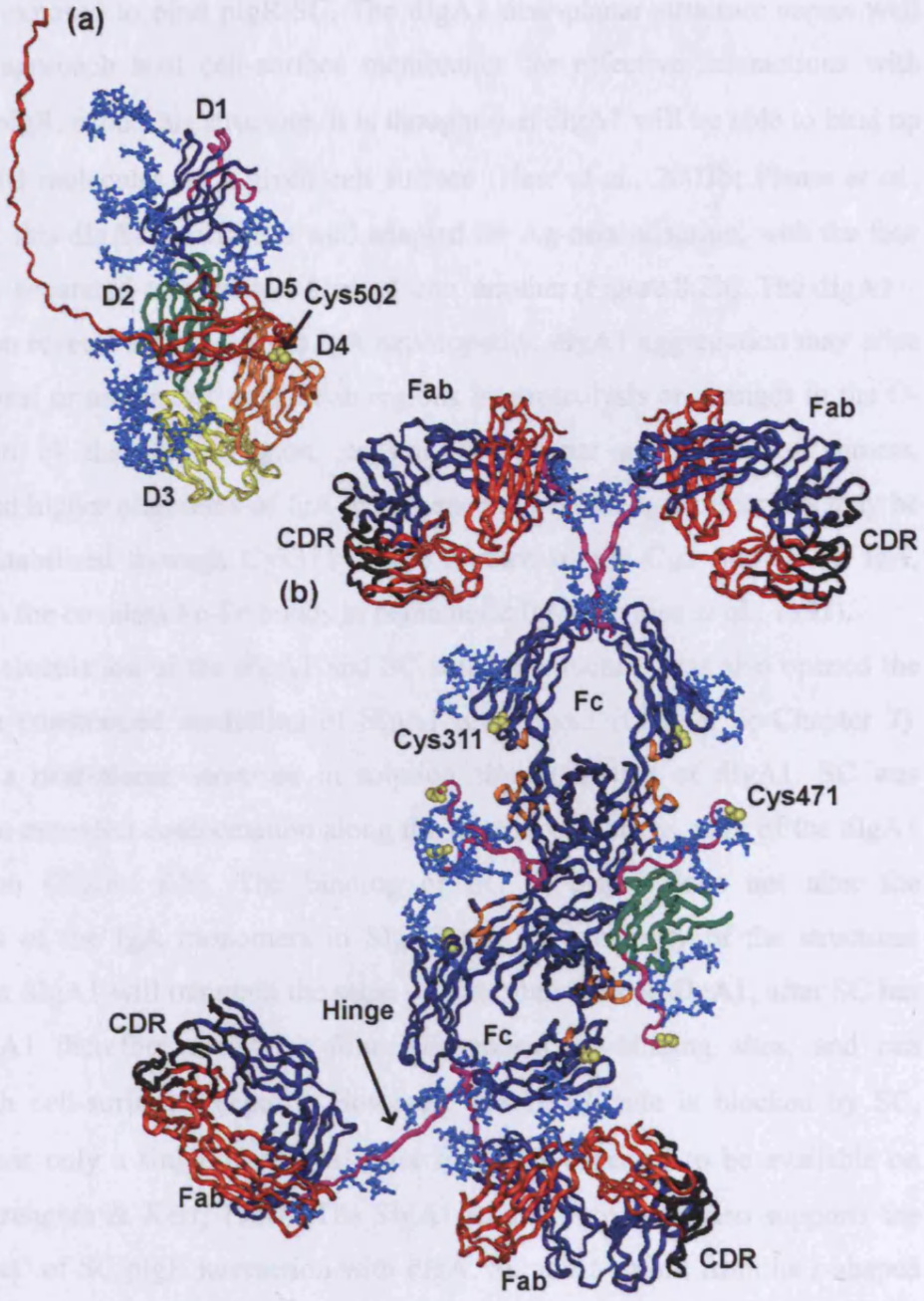

Figure 8.2 Solution structures of SC and dIgA1

(a) The J-shape arrangement of the five SC domains D1-D5 are shown, which are coloured from blue (N-terminus) to red (C-terminus). The carbohydrate moieties (light blue), the CDR-like regions (pink) and Cys502 (yellow) are highlighted. (b) The near-planar arrangement of the two IgA1 monomers within dIgA1 is shown. Heavy chains (blue), light chains (red), J chain (green), SC (black), CDRs (black), carbohydrate moieties (light blue) and the hinge and tailpiece (pink) are shown. Cys311 and Cys471 are highlighted in yellow and dIgA1 residues involved in FcαRI are shown in orange.

sufficiently exposed to bind pIgR/SC. The dIgA1 near-planar structure seems well adapted to approach host cell-surface membranes for effective interactions with FcαRI and pIgR. From this structure, it is thought that dIgA1 will be able to bind up to two FcαRI molecules on a given cell surface (Herr *et al.*, 2003b; Pleass *et al.*, 1999). Also, this dIgA1 structure is well adapted for Ag-neutralisation, with the four Fab regions separated and independent of one another (Figure 8.2b). The dIgA1 structure also reveals an insight into IgA nephropathy. dIgA1 aggregation may arise due to removal or movement of the Fab regions by proteolysis or changes in the O-glycosylation of the hinge region, causing near-planar assemblies of trimers, tetramers and higher oligomers of IgA to be generated. These IgA oligomers may be covalently stabilised through Cys311 at the surface of the C_H2 domain in IgA, analogous to the covalent Fc-Fc bonds in pentameric IgM (Perkins *et al.*, 1991).

The elucidation of the dIgA1 and SC solution structures has also opened the way for the constrained modelling of SIgA1 and SIgA2 (Chapter 6; Chapter 7). SIgA1 has a near-planar structure in solution similar to that of dIgA1. SC was located in an extended conformation along the outermost, convex edge of the dIgA1 Fc-Fc region (Figure 8.3). The binding of SC to dIgA1 does not alter the arrangement of the IgA monomers in SIgA1 and the similarity of the structures suggests that SIgA1 will maintain the same effector functions as dIgA1, after SC has bound. SIgA1 therefore maintains four independent Ag-binding sites, and can interact with cell-surface receptors. However, one FcαRI site is blocked by SC, reasoning that only a single binding site for FcαRI is expected to be available on SIgA1 (Mazengera & Kerr, 1990). The SIgA1 solution structure also supports the 'zipper effect' of SC/pIgR interaction with dIgA. SC must unfold from its J-shaped structure when it is free in solution or membrane bound as pIgR as hypothesised (Figure 8.2a). The SIgA1 solution structure provides an explanation for the mutual protection effect of dIgA1 and SC within SIgA1 (Beale, 1985; Crottet & Corthésy, 1998; Almogren *et al.*, 2003). The likely cleavage sites at the loop regions in the Fc and SC domains along the convex edge of the Fc-Fc region are masked in SIgA and will become less accessible after SIgA formation. Proteolysis at the concave edge of the Fc-Fc region may be reduced due to limited protease access, mediated by the two Fab regions above and below it, and the eight carbohydrate moieties in SIgA1.

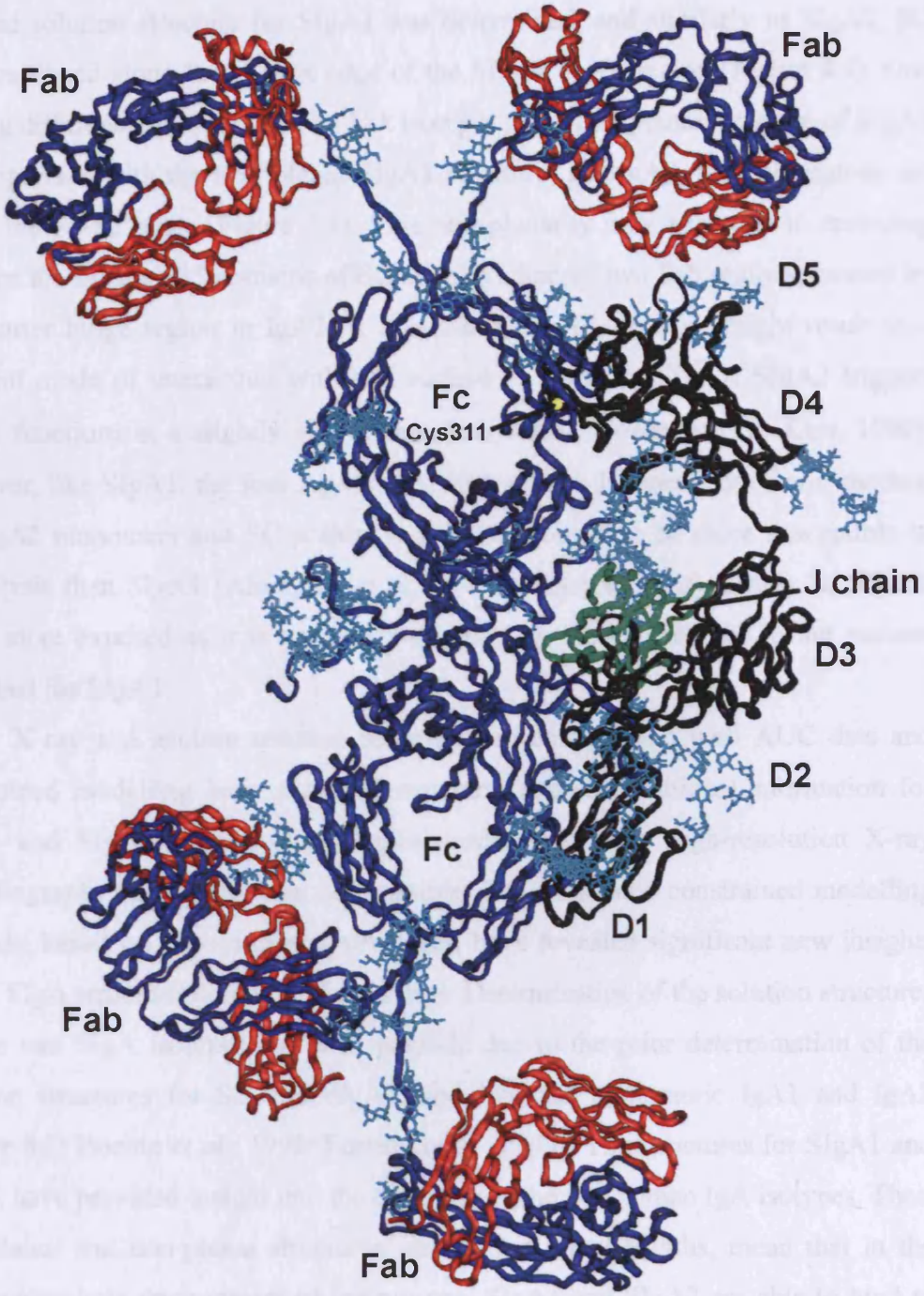

Figure 8.3 Solution structure of SIgA1

The position of SC along the convex Fc-Fc edge in SIgA1 is shown. The heavy chains (blue), light chains (red) and N-linked and O-linked oligosaccharides (cyan) are shown. The J chain (green) and SC (black) are shown. Cys311 is highlighted in yellow.

The solution structure for SIgA2 was determined, and similarly to SIgA1, SC was positioned along the convex edge of the SIgA2 Fc-Fc region (Figure 8.4). One striking difference between the two IgA isotypes is the non-planar structure of SIgA2 in comparison with the near-planar SIgA1 structure. In SIgA2, the Fab regions are not in the Fc-Fc plane (Figure 8.4). This non-planarity may arise due to crowding between the D1 and D5 domains of SC and the adjacent two Fab regions, caused by the shorter hinge region in IgA2. A non-planar SIgA2 structure might result in a different mode of interaction with cell surface receptors to SIgA1. SIgA2 triggers FcαRI functions at a slightly slower rate than SIgA1 (Mazengera & Kerr, 1990). However, like SIgA1, the four Ag-binding sites remain independent of one another. The IgA2 monomers and SC within SIgA2 are thought to be more susceptible to proteolysis than SIgA1 (Almogren *et al.*, 2003; Almogren *et al.*, 2006). In SIgA2, SC is more exposed as it is no longer shielded by the Fab regions in the manner described for SIgA1.

X-ray and neutron solution scattering, in combination with AUC data and constrained modelling have provided structural and compositional information for SIgA1 and SIgA2 in near-physiological conditions when high-resolution X-ray crystallography and NMR were not possible. The automated constrained modelling methods, based on known crystal structures, have revealed significant new insights on the SIgA structure-function relationships. Determination of the solution structures for the two SIgA isotypes has been possible due to the prior determination of the solution structures for SC and dIgA1, and likewise monomeric IgA1 and IgA2 (Figure 8.2; Boehm *et al.*, 1999; Furtado *et al.*, 2004). The structures for SIgA1 and SIgA2 have provided insight into the existence of the two human IgA isotypes. Their near-planar and non-planar structures, and varied hinge lengths, mean that in the harsh proteolytic environment of the mucosa, SIgA1 and SIgA2 are able to bind to different epitopes in a range of positions and orientations, therefore increasing the efficiency of the mucosal immune response.

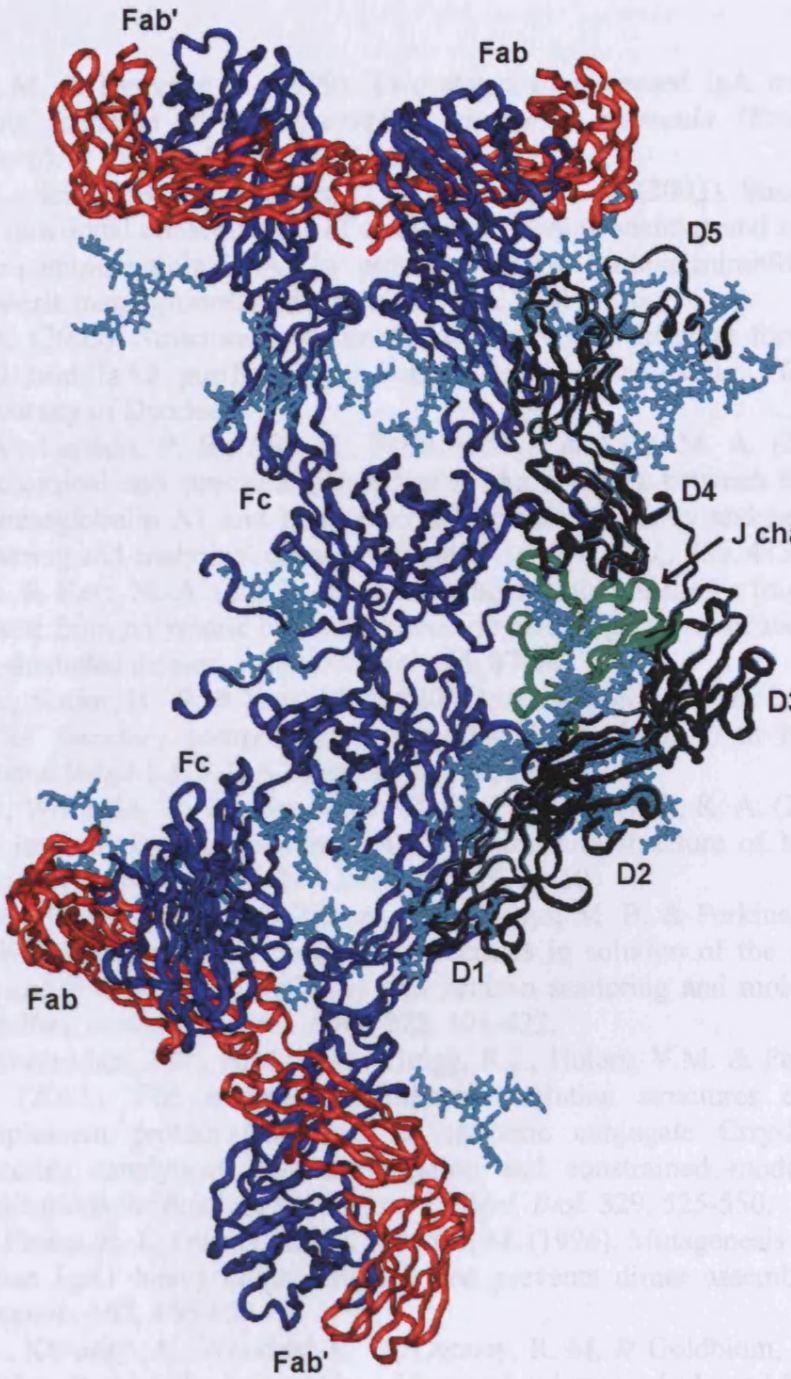

Figure 8.4 Solution structure of SIgA2

The non-planar arrangement of the IgA2 monomers within SIgA2 is shown. One Fab region on each IgA2 monomer projects into the page (Fab) and the other projects outwards (Fab'). Heavy chains (blue), light chains (red), J chain (green) and N-linked oligosaccharides (cyan) are shown. SC (black) is shown to be extended along the convex Fc-Fc SIgA2 face.

References

Adamski, F. M. & Demmer, J. (1999). Two stages of increased IgA transfer during lactation in the marsupial, *tricosurus vulpecula* (Brushtail possum). *J. Immunol.*, **162**, 6009-6015.

Almogren, A., Senior, B. W., Loomes, L. M. & Kerr, M. A. (2003). Structural and functional consequences of cleavage of human secreyrt and human serum immunoglobulin A1 by proteinases from Proteus mirabilis and Neisseria meningitides. *Infec. Immunity*, **71**, 3349-3356.

Almogren, A. (2003). Structural and functional analysis of different forms of IgA1 and IgA2 purified from human serum or colostrum. Thesis, University of Dundee.

Almogren, A., Furtado, P. B., Sun, Z., Perkins, S. J. & Kerr, M. A. (2006). Biochemical and structural properties of the complex between human immunoglobulin A1 and human serum albumin by X-ray and neutron scattering and analytical ultracentrifugation. *J. Mol. Biol.*, **356**, 413-431.

Almogren, A. & Kerr, M. A. (2007). Irreversible aggregation of the Fc fragment derived from polymeric but not monomeric serum IgA1- implications in IgA-mediated disease. *Mol. Immunol.*, **45**, 87-94.

Almogren, A., Senior, B. W. & Kerr, M. A. (2007). A comparison of the binding site of secretory component to immunoglobulin A (IgA) in human colostral S-IgA1 & S-IgA2. *Immunol.*, **120**, 273-280.

Arnold, J. N., Wormald, M. R., Sim, R. B., Rudd, P. M. & Dwek, R. A. (2007). The impact of glycosylation on the function and structure of human immunoglobulins. *Annu. Rev. Immunol.*, **25**, 21-50.

Ashton, A. W., Boehm, M. K., Gallimore, J. R., Pepys, M. B. & Perkins, S. J. (1997). Pentameric and decameric structures in solution of the serum amyloid P component by X-ray and neutron scattering and molecular modelling analyses. *J. Mol. Biol.*, **272**, 408-422.

Aslam, M., Guthridge, J.M., Hack, B.K., Quigg, R.J., Holers, V.M. & Perkins, S.J. (2003). The extended multidomain solution structures of the complement protein Crry and its chimeric conjugate Crry-Ig by scattering, analytical ultracentrifugation and constrained modelling: implications for function and therapy. *J. Mol. Biol.* **329**, 525-550.

Atkin, J. D., Pleass, R. J., Owens, R. J. & Woof, J. M. (1996). Mutagenesis of the human IgA1 heavy chain tailpiece that prevents dimer assembly. *J. Immunol.*, **157**, 156-159.

Bakos, M-A., Kurosky, A., Woodard, C. S., Denney, R. M. & Goldblum, R. M. (1991a). Probing the topography of free and polymeric Ig-bound human secretory component with monoclonal antibodies. *J. Immunol.*, **146**, 162-168.

Bakos, M-A., Kurosky, A. & Goldblum, R. M. (1991b) Characterisation of a critical binding site for human polymeric Ig on secretory component. *J. Immunol.*, **147**, 3419-3426.

Bakos, M. A., Widen, S. G. & Goldblum, R. M. (1994). Expression and purification of biologically active domain I of the human polymeric immunoglobulin receptor. *Mol. Immunol.*, **31**, 165-168.

Barnes, R. M., Allan, S., Taylor-Robinson, C. H., Finn, R. & Johnson, P. M. (1990). Serum antibodies reactive with *Saccharomyces cerevisiae* in

inflammatory bowel disease: is IgA antibody a marker for Crohn's disease? *Int. Arch. Allergy Appl. Immunol.*, **92**, 9-15.

Barratt, J. & Feehally, J. (2005). IgA nephropathy. *J. Am. Soc. Nephrol.*, **16**, 2088-2097.

Barratt, J., Smith, A. C., Molyneux, K. & Feehally, J. (2007). Immunopathogensis of IgAN. *Semin. Immunopathol.*, **29**, 427-443.

Baumann, U., Wu, A., Flaherty, K. M. & McKay, D. B. (1993). Three-dimensional structure of the alkaline protease *Psedomonas aeruginosa*: a two-domain protein with a calcium binding parallel beta roll motif. *EMBO J.*, **12**, 3357-3364.

Beale, D. (1985). Differences in fragmentation between bound and unbound bovine secretory component suggest a model for its interaction with polymeric immunoglobulin. *Biochem. J.*, **229**, 759-763.

Beavil, A. J., Young, R. J., Sutton, B. J. & Perkins, S. J. (1995). Bent domain structure of recombinant human IgE-Fc in solution by X-ray and neutron scattering in conjunction with an automated curve fitting procedure. *Biochemistry*, **34**, 14449-14461.

Berger, J. & Hinglais, N. (1968). Les dépôts intercapillaries d'IgA-IgG. *J. Urol. Néphrol.*, **74**, 694-695.

Bhown, A. S., Mole, J. E. & Bennett, J. C. (1977). Primary structure of human J chain: isolation and characterisation of tryptic and chymotryptic peptides of human J chain. *Biochem.*, **16**, 3501-3506.

Bjork, I. & Lindh, E. (1974). Gross conformation of human secretory immunoglobulin A and its component parts. *Eur. J. Biochem.*, **45**, 135-145.

Bloth, B. & Svehag, S-E. (1971). Further studies on the ultrastructure of dimeric IgA of human origin. *J. Exp. Med.*, 1035-1042.

Boehm, M. K., Woof, J. M., Kerr, M. A. & Perkins, S. J. (1999). The Fab and Fc fragments of IgA1 exhibit a different arrangement from that in IgG: A study by X-ray and neutron solution scattering and homology modelling. *J. Mol. Biol.*, **286**, 1421-1447.

Boehm, M. K. & Perkins, S. J. (2000). Structural models for carcinoembryonic antigen and its complex with the single-chain Fv antibody molecule MFE23. *FEBS Lett.* **475**, 11-16.

Boren, T., Falk, P., Roth, K. A., Larson, G. & Normark, S. (1993). Attachment of *Helicobacter pylori* to human gastric epithelium mediated by blood group antigens. *Science*, **262**, 1892-1895.

Bork, P., Holm, L. & Sander, C. (1994). The immunoglobulin fold. *J. Mol. Biol.*, **242**, 309-320.

Braathen, R., Sandvik, A., Berntzen, G., Hammerschmidt, S., Flackstein, B., Sandlie, I., Brandtzaeg, P., Johansen, F-E. & Lauvrak, V. (2006). Identification of a polymeric Ig receptor binding phage-displayed peptide that exploits epithelial transcytosis without dimeric IgA competition. *J. Biol. Chem.*, **281**, 7075-7081.

Braathen, R., Hohman, V. S., Brandtzaeg, P. & Johansen, F-E. (2007). Secretory antibody formation: conserved binding between J chain and polymeric Ig receptor from humans and amphibians. *J. Immunol.* **178**, 1589-1597.

Brandon, C. & Tooze, J. (1999). Introduction to protein structure. 2nd Edition. Garland Publishing, New York.

Brandtzaeg, P. (1970). Human secretory immunoglobulins. Thesis, Universitetsforlaget, Oslo.

Brandtzaeg, P. (1974). Mucosal and glandular distribution of immunoglobulin components: differential localization of free and bound SC in secretory epithelial cells. *J. Immunol.*, **112**, 1553-1559.

Brandtzaeg, P. (1975). Blocking effects of J chain and J chain-antibody on the binding of SC to human IgA and IgM. *Scand. J. Immunol.*, 837-842.

Brandtzaeg, P. (1983). The secretory immune system of lactating human mammary glands compared with other exocrine organs. *Ann. NY Acad. Sci.*, **409**, 353-382.

Brandtzaeg, P. & Prydz, H. (1984). Direct evidence for an itegrated function of J chain and secretory component in epithelial transport of immunoglobulins. *Nature*, 311, 71-73.

Brandtzaeg, P., Baekkevold, E. S. & Morton, H. C. (2001). From B to A the mucosal way. *Nat. Immunol.*, **2**, 1093-1094.

Brandtzaeg, P. & Johansen, F-E. (2005). Mucosal B cells: phenotypic characteristics, transcriptional regulation and homing properties. *Immunol. Rev.*, **206**, 32-63.

Brandtzaeg, P. (2007). Induction of secretory immunity and memory at mucosal surfaces. *Vaccine*, **25**, 5467-5484.

Bonner, A., Perrier, C., Corthésy, B. & Perkins, S. J. (2007). The solution structure of human secretory component and impactions for biological function. *J. Biol. Chem.*, **282**, 16969-16980.

Bonner, A., Furtado, P. B., Almogren, A., Kerr, M. A. & Perkins, S. J. (2008). Implications of the near-planar solution structure of human myeloma dimeric IgA1 for mucosal immunity and IgA nephropathy. *J. Immunol.*, **180**, 1008-1018.

Burnett, R. C., Hanly, W. C., Zhai, S. K. & Knight, K. L. (1989). The IgA heavy-chain gene family in rabbit: cloning and sequence analysis of 13 Cα genes. *EMBO J.*, **8**. 4041-4047.

Calero, M., Escribano, J., Grubb, A. & Mendez, E. (1994). Location of a novel type interpolypeptide chain linkage in the human protein HC-IgA complex and identification of a heterogeneous chromophore associated with the complex. *J. Biol. Chem.*, **269**, 384-389.

Cann, G. M., Zaritsky, A. & Elliot-Koshland, M. (1982). Primary structure of the immunoglobulin J chain from mouse. *Proc. Natl. Acad. Sci.*, **79**, 6656-6660.

Cannon, J. P., Haire, R. N. & Litman, G. W. (2002). Identification of diversified genes that contain immunoglobulin-like variable regions in a protochordate, *Nat. Immunol.*, **3**, 1200-1207.

Chacko, S. Padlan, E. A., Portolano, S., Mclachlan, S. M. & Rapoport, B. (1996). Structural studies of human antibodies. *J. Biol. Chem.*, **271**, 12191-12198.

Childers, N. K., Bruce, M. G. & McGhee, J. R. (1989). Molecular mechanisms of immunoglobulin A defence. *Annu. Rev. Microbiol.*, **43**, 503-536.

Chintalacharuvu, K. R., Raines, M. & Morrison, S. L. (1994). Divergence of human α-chain constant region gene sequences. A novel recombinant α2 gene. *J. Immunol.* **152**, 5299-5304.

Cleveland, M. G., Bakos, M-A., Pyron, D. L., Rajaraman, S. & Goldblum, R. M. (1991). Characterisation of secretory component in amniotic fluid. I dentification of new forms of secretory IgA. *J. Immunol.*, **147**, 181-188.

Collin, P., Maki, M., Keyrilainen, O., Hallstrom, O., Reunala, T. & pasternack, A. (1992). Selective IgA deficiency and celiac disease. *Scand. J. Gastroenterol.*, **27**, 367-371.

Coppo, R & Amore, A. (2004). Aberrant glycosylation in IgA nephropathy. *Kid. Int.*, **64**, 1544-1547.

Corthésy, B., Kaufmann, M., Phalipon, A., Peitsch, M., Neutram M. R. & Kraehenbuhl, J-P. (1996). A pathogen-specific epitope inserted into recombinant secretory immunoglobulin A is immunogenic by the oral route. *J. Biol. Chem.*, **271**, 33670-33677.

Corthésy, B. (2003). Recombinant secretory immunoglobulin A in passive immunotherapy: Linking immunology and biotechnology. *Current Pharmac. Biotech.*, **4**, 51-67.

Corthésy, B. (2007). Roundtrip ticket for secretory IgA: role in mucosal homeostasis. *J. Immunol.*, **178**, 27-32.

Coyne, R. S., Siebrecht, M., Peitsch, M. C. & Casanova, J. E. (1994). Mutational analysis of polymeric immunoglobulin receptor/ligand interaction. *J. Biol. Chem.*, **269**, 31620-31625.

Crottet, P. & Corthesy, B. (1998). Secretory component delays the conversion of secretory IgA into antigen-binding component $F(ab')_2$: a possible implication for mucosal defence. *American Ass. Immunologists*. 5445-5452.

Crottet, P., Cottet, S. & Corthesy, B. (1999).Expression, purification and biochemical characterisation of recombinant murine secretory component: a novel tool in mucosal immunity. *Biochem J.*, **341**, 299-306.

Crottet, P. & Corthésy, B. (1999). Mapping the interaction between murine IgA and murine secretory component carrying epitope substitutions reveals a role of domains II and III in covalent binding. *J. Biol. Chem.* **274,** 31456-31462.

Cunningham-Rundles, C. (2001). Physiology of IgA and IgA deficiency. *J. Clin. Immunol.*, **21**, 303-309.

Dallas, S. D. & Rolfe, R. D. (1998). Binding of Clostridium difficile toxin A to human milk secretory component. *J. Med. Microbiol.*, **47**, 879-888.

Davis, A. C., Roux, K. H. & Shulman, M. J. (1988). On the structure of polymeric IgM. *Eur. J. Immunol.*, **18**, 1001-1008.

Dechant, M. & Valerius, T. (2001). IgA antibodies for cancer therapy. *Crit. Rev. Oncol./Haematol.*, **39**, 69-77.

Decot, V., Woerly, G., Loyens, M., Loiseau, S., Quatannens, B., Capron, M. & Dombrowicz. (2005). Heterogeneity of expression of IgA receptors by human, mouse and rat eosinophils. *J. Immunol.*, **174**, 628-635.

Deisenhofer, J. (1981). Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from *Staphylococcus aureus* at 2.9 and 2.8Å resolution. *Biochem.*, **20**, 2361-2370.

Delacroix, D. L. & Vaerman, J. P. (1982). Secretory component (SC): preferential binding to heavy (greater than 11S) IgA polymers and IgM in serum, in contrast to predominance of 11S and free SC forms in secretions. *Clin. Exp. Immunol.*, **49**, 717-724.

Deviere J., Content J., Denys C., Vandenbussche P., Le Moine O., Schandene, L., Vaerman J. P. & Dupont E. (1992). Immunoglobulin A and

interleukin 6 form a positive secretory feedback loop: a study of normal subjects and alcoholic cirrhotics. *Gastroenterol.*, **103**,1296-1301.

Dourmashkin, R. P., Virella, G. & Parkhouse, R. M. E. (1971). Electron microscopy of human and mouse myeloma serum IgA. *J. Mol. Biol.*, **56**, 207-208.

Elm, C., Braathen, R., Bergmann, S., Frank, R., Vaerman, J-P., Kaetzel, C. S., Chhatwal, G. S., Johansen, F-E. & Hammerscmidt, S. (2004). Ectodomain 3 and 4 of human polymeric immunoglobulin receptor (hpIgR) mediate invasion of *Streptococcus pneumoniae* into the epithelium. *J. Biol. Chem.*, **279**, 6296-6304.

Fallgreen-Gebauer, E., Gebauer, W., Bastian, A., Kratzin, H. D., Eiffert, H., Zimmermann, B., Karas, M. & Hilschmann, N. (1993). A covalent linkage of secretory component to IgA. Structure of SIgA. *Biol Chem Hoppe Seyler*, **374**, 1023-1028.

Feinstein, A., Munn, E. A. & Richardson, N. E. (1971). The three-dimensional conformation of M and A globulin molecules. *Ann. N. Y. Acad. Sci.*, **190**, 104-121.

Feehally, J. & Allen, A. (1999). Structural features of IgA molecules which contribute to IgA nephropathy. *J. Nephrol.*, **12**, 59-65.

Ferkol, T., Perales, J C., Eckman, E., Kaetzel, C. S., Hanson, R. W. & Davis, P. B. (1995). Gene transfer into the airway epithelium of animals by targeting the polymeric immunoglobulin receptor. *J. Clin. Invest.*, **95**, 493-502.

Flanagan, J. G., Lefranc, M-P. & Rabbitts, T. H. (1984). Mechanisms of divergence and convergence of the human immunoglobulin $\alpha 1$ and $\alpha 2$ constant region gene sequences. *Cell*, **36**, 681-688.

Frutiger, S., Hughes, G.J., Hanly, C., Kingzette, M., Jaton, J-C. (1986) The amino-terminal domain of rabbit secretory component is responsible for noncovalent binding to immunoglobulin A dimers. *J. Biol. Chem.*, **261**, 16673-16681.

Frutiger, S., Hughes, G. J., Hanly, W. C., Kingzette, M. & Jaton, J-C. (1987). Structural variability of rabbit secretory component. *J. Biol. Chem.*, **262**, 10463- 10469.

Frutiger, S., Hughes, G. J., Paquet, N., Luthy, R. & Jaton, J-C. (1992). Disulphide bond assignment in human J chain and its covalent pairing with immunoglobulin M. *Biochem.*, **31**, 12643-12647.

Fry, L. (2002). Dermatitis herpetiformis: problems, progress and prospects. *Eur. J. Dertmatol.*, **12**, 523-531.

Furtado, P. B., Whitty, P. W., Robertson, A., Eaton, J. T., Almogren, A., Kerr, M. A., Woof, J. M. & Perkins, S. J. (2004). Solution structure determination of human IgA2 by X-ray and neutron scattering and analytical ultracentrifugation and constrained modelling: a comparison with human IgA1. *J. Mol. Biol.*, **338**, 921-941.

Gala, F. A., & Morrison, S. L. (2002). The role of constant region carbohydrate in the assembly and secretion of human IgD and IgA1. *J. Biol. Chem.*, **277**, 29005-29011.

Garcia de la Torre, J., Huertas, M. L. & Carrasco, B. (2000). Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.*, **78**, 719-730.

Garcia-Pardo, A., Lamm, M. E., Plaut, A. G. & Frangoine, B. (1981). J chain is covalently bound to both monomer subunits in human secretory IgA. *J. Biol. Chem.*, **256**, 11734-11738.

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D. & Bairoch A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook. (Walker, J. M. ed) Humana Press.

Geisberger, R., Lamers, M. & Achatz, G. (2006). The riddle of the dual expression of IgM and IgD. *Immunol.*, **118**, 429-437.

Gilbert, H. E., Aslam, M., Guthridge, J. M., Holers, V. M. & Perkins, S. J. (2006). Extended flexible linker structures in the complement chimaeric conjugate CR2-Ig by scattering, analytical ultracentrifugation and constrained modelling: implications for function and therapy. *J. Mol. Biol.*, **356**, 397-412.

Glatter, O. and Kratky, O., (1982) Small Angle X-ray Scattering. Academic Press, New York.

Goldsby, R. A., Kindt, R. A., Osborne, B. A. & Kuby, J. (2002) Antibodies: structure and function. In Immunology 5th Edition. W. H. Freeman & Company, New York, pp 76-104.

Gorter, A., Hiemstra, P. S., Leijh, P. C. J., van de Sluys, M. E., van de Barselaar, M. T., van Es, L. A & Daha, M. R. (1987). IgA- and secretory IgA-opsonised *S. aureus* induce a respiratory burst and phagocytosis by polymorphonuclear leucocytes. *Immunol.*, **61**, 303-309.

Grey, H. M., Abel, C. A., Yount, W. J. & Kinkel, H. G. (1968). A subclass of human γα-globulins (γA2) which lacks the disulphide bonds linking heavy and light chains. *J. Exp. Med.*, **128**, 1223-1236.

Grossetette B., Viard J. P., Lehuen A., Bach J. F. & Monteiro C. (1995). Impaired Fc alpha receptor expression is linked to increased immunoglobulin A levels and disease progression in HIV-1-infected patients. *AIDS*, **9**, 229-234.

Grubb, A. O., Lopez, C., Tejler, L. & Mendez, E. (1983). Isolation of human complex-forming glycoprotein heterogeneous in charge (protein HC) and its IgA complex from plasma. *J. Biol. Chem.*, **258**, 14698-14707.

Guinier, A. & Fournet, G. (1955). Small-angle scattering of X-rays. Wiley, New York.

Gulle, H., Samstag, A., Eibl, M. M. & Wolf, H. M. (1998). Physical and functional association of Fc alpha R protein tyrosine kinase Lyn. *Blood*, **91**, 383-391.

Halpern, M. S. & Koshland. (1970). Novel subunit in secretory IgA. *Nature*, **228**, 1276-1278.

Hamada, K., Hata, Y., Katsuya, Y., Hiramatsu, Fujiwara, T. & Katsube, Y. (1996). Crystal structure of *Serratia* protease, a zinc-dependent proteinase from *Serratia* sp. E-15, containing a beta-sheet coil motif at 2.0 Å resolution. *J. Biochem.*, **119**, 844-851.

Hamburger, A. E., West Jr. A. P. & Bjorkman, P. J. (2004). Crystal structure of a polymeric immunoglobulin binding fragment of the human polymeric immunoglobulin receptor. *Structure*, **12**, 1925-1935.

Hanson L. A. (1983). Selective IgA-deficiency. In Primary and secondary immunodeficiency disorders. (Chandra R. K. ed). Churchill Livingstone, New York, pp 62–84.

Heenan R. K., Penfold J. & King, S. M. (1997). SANS at pulsed neutron sources: present and future prospects. *J. Appl. Cryst.*, **30,** 1140-1147.

Heenan, R. K., King, S. M., Osborn, R. & Stanley, H. B. (1989). COLLETTE Users Guide. Internal publication RAL-89-128, Rutherford Appleton Laboratory, Didcot, U.K.

Heimstra, P. S., Biewenga, J., Gorter, A., Stuurman, M. E., Faber, A., van Es, L. A. & Daha, M. R. (1988). Activation of complement by human serum IgA, secretory IgA and IgA1 fragments. *Mol. Immunol.*, **25**, 527-533.

Hendrickson, B. A., Rindisbacher, L., Corthesy, B., Kendall, D., Waltz, D. A., Neutra, M. R. & Seidman, J. G. (1996). Lack of association of secretory component with IgA in J chain deficient mice. *J. Immunol.*, **157**, 750-754.

Heremans, J. F. (1974). Immunoglobulin A. In 'The antigens' volume II. Sela, M. (ed) Academic Press, New York, 365-522.

Herr, A. B., White, C. L., Milburn, C., Wu, C. & Bjorkman, P. J. (2003a). Bivalent binding of IgA1 to FcαRI suggests a mechanisms for cytokine activation of IgA phagocytosis. *J. Mol. Biol.*, 327, 645-657.

Herr, A. B., Ballister, E. R. & Bjorkman, P. J. (2003b). Insights into IgA-mediated immune responses from the crystal structures of human FcαRI and its complex with IgA1-Fc. *Nature*, **423**, 614-620.

Hexham, J. M., White, K. D., Carayannopoulos, L. N., Mandecki, W., Brisette, R, Yang, Y-S. & Capra, J. D. (1999). A human immunoglobulin (Ig)A Cα3 domain motif directs polymeric Ig receptor-mediated secretion. *J. Exp. Med.*, **189**, 747-751.

Hohman, V. S., Stewart, S. E., Willet, C. E. & Steiner, L. A. (1997). The sequence of the J chain in the amphibian, *Xenopus laevis*. *Mol. Immunol.*, **34**, 995-1002.

Hohman, V. S. & Steiner, L. A. (2000). The sequence of J chain in an amphibian, *Rana catesbeiana*. *Immunogenetics*, **51**, 587-590.

Hohman, V. S., Stewart, S. E., Stewart, S. E., Rumfelt, L. L., Greenberg, D. W., Avila, D. W., Flajnik, M. F. and Steiner, L. A. (2003). J chain in the nurse shark: implications for function in a lower vertebrate. *J. Immunol.*, **170**, 6016-6023.

Huang, T. C., Toraya, H., Blanton, T. N. & Wu, Y. (1986). X-ray powder diffraction analysis of silver behenate, a possible low-angle diffraction standard. *J. Appl. Cryst.*, **26**, 180-184.

Hughes, G. J., Frutiger, F-E., Paquet, N & Jaton, J-C. (1990). The amino acid sequence of rabbit J chain in secretory IgA. *Biochem. J.*, **271**, 641-647.

Hughes, G. J., Frutiger, S., Savoy, L-A., Reason, A. J., Morris, H. R., Jaton, J-C. (1997). Human free secretory component is composed of the first 585 amino acid residues of the polymeric immunoglobulin receptor. *FEBS Letters*, **410**, 443-446.

Hu, Y., Sun, Z., Eaton, J. T., Bouloux, P. M. G. & Perkins, S. J. (2005). Extended and flexible domain solution structure of the extracellular matrix protein anosmin-1 by X-ray scattering, analytical ultracentrifuation and constrained modelling. *J. Mol. Biol.*, **350**, 553-70.

Iwata, A., Iwase, Oguta, Y., Takahashi, T., Matsumoto, N., Yoshida, T., Kamei, N., Kobiyashi, K., Mestecky, J. & Moro, I. (2002). Cloning and expression of the turtle (*Trachemys scripta*) immunoglobulin joining (J)-chain cDNA. *Immunogenetics*, **54**, 513-519.

Janeway, C. (2004). Immunobiology 6th Edition. Garland Publishing, New York.

Johansen, F-E., Natvig Norderhaug, I., Roe, M., Sandlie, I. & Brandtzaeg, P. (1999). Recombinant expression of polymeric IgA: incorporation of J chain and secretory component of human origin. *Eur. J. Immunol.*, **29**, 1701-1708.

Johansen, F-E., Braathen, R. & Brandtzaeg, P. (2001). The J chain is essential for polymeric Ig receptor-mediated epithelial transport of IgA. *J. Immunol.*, **167**, 5185-5192.

Kabat, E. A., Wu, T. T. & Bilofsky, H. (1976). Attempts to locate residues in complementarity-determining regions of antibody combining sites that make contact with antigen. *Proc. Nat. Ac. Sci. USA*, **73**, 617-619.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577-2637.

Kerr, M. A. (1990). The structure and function of human IgA. *Biochem. J.*, **271**, 285-296.

Kerr, M. A & Woof, J, M. (1999). Fcα receptors. In Mucosal immunology, 213-223.

Kratky, O., Pilz, I. & Scmitz, P. J. (1966). Absolute intensity measurement of small angle X-ray scattering by means of a standard sample. *J. Colloid Interface Sci.*, **21**, 24-34.

Krugman, S., Pleass, R. J., Atkin, J. D. & Woof, J. M. (1997). Strutcural requirements for assembly of IgA probed by site-directed mutagenesis of J chain and cysteine residues of the α-chain CH2 domain. *J. Immunol.*, **159**, 244-249.

Kulseth, M. A. & Rogne, S. (1994). Cloning and characterisation of the bovine immunoglobulin J chain cDNA and its promoter region. *DNA cell Biol.*, **13**, 37-42.

Lamkhioued, B., Gounni, A. S., Gruart, V., Pierce, A., Capron, A. & Capron, M. (1995). Human eosinophils express a receptor for secretory component. Role in secretory IgA-dependent activation. *Eur. J. Immunol.*, **25**, 117-125.

Lamm, M. E. (1997). Interaction of antigens and antibodies at mucosal surfaces. *Annu. Rev. Microbiol.*, **51**, 311-340.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, **26**, 283-291.

Laue, T. M., Shah, B. D., Ridgeway, T. M., and Pelletier, S. L. (1992). Computer-aided interpretation of analytical sedimentation data for proteins. *In* Analytical Ultracentrifugation in Biochemistry and Polymer Science (Harding, S. E., Rowe, A. J. and Horton, J. C., eds) The Royal Society of Chemistry, Cambridge, U.K., 90-125.

Launay, P., Grossetête, B., Arcos-Fajardo, M., Gaudin, E., Torres, S. P., Beaudoin, L., de Serre, N., P-M., Lehuen, A. & Monteiro, R. C. (1999). FcαR (CD89) mediates the development of immunoglobulin A nephropathy (Berger's disease): evidence for pathogenic soluble-receptor-IgA complexes in patients and CD89 transgenic mice. *J. Exp. Med.*, **11**, 1999-2009.

Lesk, A. M. & Chothia, C. (1982). Evolution of proteins formed by β-sheets. The core of the immunoglobulin domain. *J. Mol. Biol.* **160**, 325-342.

Lindh, E. (1975). Increased resistance of immunoglobulin A dimers to proteolytic degradation after binding of secretory component. *J. Immunol.*, **114**, 284-286.

Loomes, L. M., Stewart, W. W., Mazengera, R. L., Senior, B. W. & Kerr, M. A. (1991). Purification and characterization of human immunoglobulin IgA1 and IgA2 isotypes from serum. *J. Immunol. Methods,* **141**, 209-218.

Lu, L., Lamm, M. E., Li, H., Corthésy, B. & Zhang, J-R. (2003). The human polymeric immunoglobulin receptor binds to *Streptococcus pneumoniae* via domains 3 and 4. *J. Biol. Chem.* **278**, 48178-48187.

Lullau, E., Heyes, S., Vogel, H., Marison, I., von Stockar, U., Kraehenbuhl, J-P. & Corthesy, B. (1996). Antigen binding properties of purified immunoglobulin A and reconstituted secretory immunoglobulin A antibodies. *J. Biol. Chem.* **271**, 16300-16309.

Ma, J. K. C., Hikmat, B. T., Wycoff, K., Vine, N. D., Chargelegue, D., Yu, L., Hein, M. B. & Lehner, T. (1998). Characterisation of a recombinant plant monoclonal secretory antibody and preventative immunotherapy in humans. *Nat. Med.*, **4**, 601-606.

Mach, J-P. (1970). *In vitro* combination of human and bovine free sceretroy component with IgA of various species. *Nature*, **288**, 1278-1282.

Macpherson, A. J. & Harris, N. L. (2004). Interactions between commensal intestinal bacteria and the immune system. *Nat. Rev. Immunol.*, **4**, 478-485.

McDonald, K. J., Cameron, A. J., Allen, J. M. & Jardine, A. G. (2002). Expression of Fc alpha/mu receptor by human mesangial cells: a candidate receptor for immune complex deposition in IgA nephropathy. *Biochem. Biophys. Res. Commun.*, **290**, 438-442.

Mantis, N. J., Cheung, M. C., Chinalacharuvu, K. R., rey, J., Corthesy, B. & Neutra, M. R. (2002). Selective adherence of IgA to murine Peyer's patches M cells: evidence for a novel IgA receptor. *J. Immunol.*, **169**, 1844-1851.

Matsuuchi, L. G., Cann, G. M & Koshland, M. E. (1986). Immunoglobulin J chain gene from the mouse. *Proc. Natl. Acad. Sci. USA*, **83**, 456-460.

Mattey, D. L., Nixon, N. B., Dawes, P. T., Scott, L. J. & Russell, G. I. (1997). Measurement of IgA1-α1 anti-trypsin complex in rheumatoid arthritis: A question of specificity? *Ann. Rheum. Dis.*, **56**, 439-440.

Mattu, T. S., Pleass, R. J., Willis, A. C., Kilian, M. & Wormald, M. R. (1998). The glycosylation and structure of human serum IgA, Fab and Fc regions and the role of N-glycosylation on Fcα receptor interactions. *J. Biol. Chem.*, **273**, 2260-2272.

Max, E. E. & Korsmeyer, S. J. (1985). Human J chain gene: structure and expression in B lymphoid cells. *J. Exp. Med.*, **161**, 832-849.

Mazengera, R. L. & Kerr, M. A. (1990). The specificity of the IgA receptor purified from human neutrophils. *Biochem. J.*, **272**, 159-165.

Mestecky, J., Schrohrnloher, R. E., Kulhavy, R., Wright, G. P. & Tomana, M. (1974). Site of J chain attachement to human polymeric IgA. *Nature*, 249, 550-552.

Mestecky, J., Hammack, W. J., Kulhavy, R., Wright, G. P. & Tomana, M. (1977). Properties of IgA myeloma proteins isolated from sera of

patients with the hyperviscosity syndrome. *J. Lab. Clin. Med.*, **89**, 919-927.

Mestecky, J. & McGhee, J. R. (1987). Immunoglobulin A: molecular and cellular interactions involved in IgA biosynthesis and immune response. *Adv. Immunol.*, **40**, 153-245.

Mestecky, J., Russell, M. W. & Elson, C. O. (1999). Intestinal IgA: novel views on its function in the defence of the largest mucosal surface. *Gut*, **44**, 2-5.

Mestecky, J., Moro, I., Kerr, M. A. & Woof, J. M. (2005). Mucosal Immunoglobulins. In Mucosal Immunology 3rd Edition. Mestecky, J. R., Bienenstock, J., Lamm, M. E., Mayer, L., McGhee, J. R. & Strober, W. (eds). Academic Press, San Diego, 153–181.

Morrison, S. L. & Koshland, M. E. (1972). Characterisation of the J chain from polymeric immunoglobulins. *Proc. Nat. Acad. Sci. USA*, **69**, 124-128.

Monteiro, R. C., Moura, I. C., Launay, P., Tsuge, T., Haddad, E., Benhamou, M., Cooper, M. C. & Arcos-Fajardo, M. (2002). Pathogenic significance of IgA receptor interactions in IgA nephropathy. *Trends Mol. Med.*, **8**, 464-468.

Monterio, R. C. & van de Winkle, J. G. J. (2003). IgA Fc receptors. *Annu. Rev. Immunol.*, **21**, 177-204.

Morton, H. C., van Egmond, M. & van de Winkle, G. J. (1996). Structure and function of human IgA Fc receptors (FcαR). *Critical Rev. Immunol.*, **16**, 423-440.

Mostov, K. E. (1994). Transepithelial transport of immunoglobulins. *Annu. Rev. Immunol.*, **12**, 63-84.

Moura, I. C., Centelles, M. N., Arcos-Fajardo, M., Malheiros, D. M., Collawn, J. F., Cooper, M. D. & Monteiro, R. C. (2001). Identification of the transferring receptor as a novel immunoglobulin (Ig)A1 receptor and its enhanced expression on mesangial cells in IgA nephropathy. *J. Exp. Med.*, **194**, 417-425.

Moura, I. C., Arcos-Fajardo, M., Leroy, V., Sadaka, C., Mahlaoui, N., Lepelletier, Y., Vrtovsnik, F., Haddad, E., Benhamou, M. & Monteiro, R. C. (2005). Engagement of transferring receptor by polymeric IgA1: evidence for a positive feedback loop involving increased receptor expression and mesangial cell proliferation in IgA nephropathy. *J. Am. Soc. Nephrol.*, **16**, 2667-2676.

Munn, E. A., Feinstein, A. & Munro, A. J. (1971). Electron microscopy examination of free IgA molecules and of their complexes with antigen. *Nature*, **231**, 527-529.

Murakami, M. & Honjo, T. (1995). Invovement of B-1 cells in mucosal immunity and autoimmunity. *Immunol. Today*, **16**, 534-539.

Nagler-Anderson, C. (2001). Man the barrier! Strategic defences in the intestinal mucosa. *Nature Rev. Immunol.*, **1**, 59-67.

Narayanan, T., Diat, O. & Bosecke, P. (2001). SAXS and USAXS on the high brilliance beamline at the ESRF. *Nucl. Instrum. Methods Phys. Res.*, **467-468**, 1005-1009.

Nezlin, R. (1998). The immunoglobulins: Structure and function. *Academic Press*, N.Y.

Norderhaug, I. N, Johansen, F.-E., Krajci, P. & Brandtzaeg, P. (1999). Domain deletions in the human polymeric Ig receptor disclose differences

between its dimeric IgA and pentameric IgM interaction. *Eur J. Immunol.*, **29**, 3401-3409.

Ondo, K., Ohi, H., Tamano, M., Ohsawa, I., Wakabayashi, M., Horikoshi, S., Fujita, T. & Tomino, Y. (2007). Hypercomplementemia in adult patients with IgA nephropathy. *J. Clin. Lab. Anal.*, **21**, 77-84.

Oortwijn, B. D., Rastaldi, M. P., Roos, A., Mattinzoli, D., Daha, M. R. & van Kooten, C. (2007). Demonstration of secretory IgA in kidneys of patients with IgA nephropathy. *Nephrol. Dial. Transplant.*, **22**, 3191-3195.

Oortwijn, B. D., Roos, A., van der Boog, P. J. M., Klar-Mohamad, N., van Remoortere, A., Deelder, A. M., Daha, M. R. & van Kooten, C. (2007). Monomeric and polymeric IgA show a similar association with the myeloid FcαRI/CD89. *Mol. Immunol.*, **44**, 966-973.

Orengo, C. A., Jones, D. T. & Thornton, J. M. (1994). Protein superfamilies & domain superfolds. *Nature*, **372**, 631-634.

Otten H. G., Daha M. R., van der Maarl M. G., Hoogendoorn L. I., Beem E. M., de Roo, H. H. & Breedveld F. C. (1992). IgA rheumatoid factor in mucosal fluids and serum of patients with rheumatoid arthritis: immunological aspects and clinical significance. *Clin. & Exp. Immunol.*, **90**, 256-259.

Panine, P., Finet, S., Weiss, T. M. & Narayanan, T. (2006). Probing fast kinetcs in complex fluids by combined rapid mixing and small-angle X-ray scattering. *Adv. Colloid. Interface Sci.*, **127**, 9-18.

Perkins, S. J. & Weiss, H. (1983). Low resolution structural studies of mitochondrial ubiquinol-cytochrome c reductase in detergent solutions by neutron scattering. *J. Mol. Biol.* **168**, 847-866.

Perkins, S. J. (1986). Protein volumes and hydration effects: the calculation of partial specific volumes, neutron scattering matchpoints and 280 nm absorption coefficients for proteins and glycoproteins from amino acid sequences. *Eur. J. Biochem.*, **157**, 169-180.

Perkins, S. J. (1988). Structural studies of proteins by high-flux X-ray and neutron solution scattering. *Biochem. J.*, **254**, 313-327.

Perkins, S. J. Nealis, A., Sutton, B. J., & Feinstein, A. (1991). Solution structure of human and mouse immunoglobulin M by synchrotron X-ray scattering and molecular graphics modelling: a possible mechanism for complement activation. *J. Mol. Biol.*, **221**, 1245-1366.

Perkins, S. J., Ashton, A. W., Boehm, M. K. & Chamberlain, D. (1998). Molecular structures from low angle X-ray and neutron scattering studies. *Int. J. Biol. Macromol.*, **22**, 1-16.

Perkins, S. J. (2001). X-ray and neutron scattering analyses of hydration shells: a molecular interpretation based on sequence predictions and modelling fits. *Biophys. Chem.*, **93**, 129-139.

Perkins, S. J., Nealis, A., Sutton, B. J. & Feinstein, A. (1991). Solution structure of human and mouse immunoglobulin M by synchrotron X-ray scattering and molecular graphics modelling: a possible mechanism for complement activation. *J. Mol. Biol.*, **221**, 1345-1366.

Perkins, S. J., Gilbert, H. E., Lee, Y. C., Sun, Z. and Furtado, P. B. (2005). Relating small angle scattering and analytical ultracentrifugation in multidomain proteins. In Modern Analytical Ultracentrifugation:

Techniques and Methods. Scott, D. J., Harding, S. E. & Rowe, A. J., (eds). Royal Society of Chemistry, London. 291-319

Perkins, S. J., Okemefuna, A. I., Fernando, A. N., Bonner, A., Gilbert, H. E. & Furtado, P. B. (2008). X-ray and neutron scattering data and their constrained molecular modelling. *Methods in cell Biol.*, **84**, 375-423.

Perkins, S. J. & Bonner, A. (2008). Structure determination of human and chimeric antibodies by solution scattering and constrained molecular modelling. *Biochem. Soc. Trans.*, **36**, 37-42.

Perrier, C., Sprenger, N. & Corthésy, B. (2006). Glycans on secretory component participate in innate protection against mucosal pathogens. *J. Biol. Chem.*, **281**, 14280-14287.

Peterson, P. A., Cunningham, B. A., Berggard, I. & Edelman, G. (1972). β_2-microglobulin – a free immunoglobulin domains. *Proc. Nat. Acad. Aci USA*, **7**, 1697-1701.

Phalipon, A., Cardona, A., Kraehenbuhl, J-P., Edelman, L., Sansonetti, P. J. & Corthésy, B. (2002). Secretory component: a new role in secretory IgA-mediated immune exclusion *in vivo*. *Immunity*, **17**, 107-115.

Phalipon, A., Corthésy, B. (2003). Novel functions of the polymeric Ig receptor: well beyond transport of immunoglobulins. *Trends Immunol.*, **24**, 55-58.

Philo, J. (2000). A method for directly fitting the time derivation of sedimentation velocity data and an alternative algorithm for calculating sedimentation coefficient distribution functions. *Anal. Biochem.*, **279**, 151-163.

Pilz, I., Kratky, O., Licht, A. & Sela, M. (1973). Shape and volume of anit-poly(D-alanyl) antibodies in the presence of tetra-D-alanine followed by small-angle scattering. *Biochem.*, **12**, 4998-5005.

Pleass, R. J., Dunlop, J. I., Anderson, C. M., Woof, J. M. (1999). Identification of residues in the CH2/CH3 domain interface of IgA essential for interaction with the human Fc alpha receptor (FcαRI) CD89. *J. Biol. Chem.*, **274**, 23508-23514.

Poljak, R. J., Amzel, L. M., Avey, H. P., Chen, B. L., Phizackerley, R. P. & Saul, F. (1973). Three-dimensional structure of the Fab' fragment of a human immunoglobulin at 2.8Å resolution. *Proc. Nat. Acad. Sci. USA*, **70**, 3305-3310.

Preud'homme, J-L., Petit, I., Barra, A., Morel, F., Lecron, J-C. & Lelièvre, E. (2000). Structural and functional properties of membrane and secreted IgD. *Mol. Immunol.*, **37**, 871-887.

Putnam, F. w., Liu, Y-S. V. & Low, T. L. K. (1979). Primary structure of a human IgA1 immunoglobulin. *J. Biol. Chem.*, **254**, 2865-2874.

Pyne, D., Ehrenstein, M. & Morris, V. (2002). The therapeutic uses of intravenous immunoglobulins in autoimmune rheumatic diseases. *Rheumatology*, **41**, 367-374.

Ralston, G. (1993). Introduction to analytical ultracentrifugation, Beckman instruments, Palo Alto.

Rhodes, G. (2000). Crystallography made crystal clear 2nd Edition. Academic Press, New York.

Rifai, A., Fadden, K., Morrison, S. J. & Chintalacharuvu, K. R. (2000). The N-glycans determine the differential blood clearance and hepatic uptake of human IgA1 and IgA2. *J. Exp. Med.*, **191**, 2171-2182.

Robins, G. & Howdle, P. D. (2005). Advances in celiac disease. *Curr. Opin. Gastroenterol.*, **21**, 152-161.

Roos, A., Bouwman, L. H., van Gijlswijk-Jassen, D. J., Faber-Krol, M. C., Stahl, G. L. & Daha, M. R. (2001). Human IgA activates the complement system via the mannan-binding lectin pathway. *J. Immunol.*, **167**, 2861-2868.

Rost, B. & Liu, J. (2003). The PredictProtein Server. *Nucleic Acids Res.*, **31**, 3300-3304.

Roux, K. H., Strelets, L., Brekke, O. H., Sandlie, I. & Michaelsen, T. E. (1998). Comparisons of the ability of human IgG3 hinge mutants, IgM, IgE and IgA2, to form small immune complexes: a role for flexibility and geometry. *J. Immuno.*, **161**, 4083-4090.

Sali, A. & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779-815.

Sato, T., can Dixhoorn, M. G., Schroeijers, W. E., van Es, L. A. & Daha, M. R. (1997). Efficient induction of apoptosis in cultured rat glomerular mesangial cells by dimeric monoclonal IgA anti-Thy-1 antibodies. *Kindney Int.*, **51**, 173-181.

Savilahti, E., Klemola, T., Carlsson, B., Mellander, L., Stenvile, M. & Hovi, T. (1988). Inadequacy of mucosal IgM antibodies in selective IgA deficiency: excretion of attenuated polio viruses is prolonged. *J. Clin. Immunol.*, **8**, 89-94.

Schaffer, A. A., Aravind, L., Madden, T. L., Shavirin, S., Spouge, J. L., Wolf, Y. I., Koonin, E. V. & Altschul, S. F. (2001). Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.*, **29**, 2994-3005.

Schuck, P. (1998). Sedimentation analysis of non-interacting and self-associating solutes using numerical solutions to the Lamn equation. *Biophys. J.* 75, 1503-512.

Scott, L. J., Evans, E. L., Dawes, P. T., Russell, G. I. & Mattey, D. L. (1998). Comparison of IgA-α1anti-trypsin levels in rheumatoid arthritis and seronegative oligoarthritis: complex formation is not associated with inflammation. *Br. J. Rheumatol.*, **37**, 398-404.

Semenyuk, A. V. & Svergun, D. I. (1991). GNOM- a program package for small-angle scattering data-processing. *J. Appl. Crystallog.* **24**, 537-540.

Senior, B. W. & Woof, J. M. (2005). The influences of hinge length and composition on the susceptibility of human IgA to cleavage by diverse bacterial IgA1 proteases. *J. Immunol.*, **174**, 7792-7799.

Shibuya, A., Sakamoto, N., Shimizu, Y., Osawa, M., Hiroyama, T., Eyre, H. J., Sutherland, G. R., Endo, Y., Fujita, T., Miyabayashi, T., Sakano, S., Tsuji, T., Nakayama, E., Phillips, J. H., lanier, L. L. and Nakauchi, H. (2000). Fc alpha/mu receptor mediates endocytosis of IgM-coated microbes. *Nat. Immunol.*, **1**, 441-446.

Smith, D. K. & Xue, H. (1997). Sequence profiles of immunoglobulin and immunoglobulin-like domains. *J. Mol. Biol.*, **274**, 530-545.

Stanworth, D. R., Lewin, I. & Crockson, R. A. (1985). Measurement of IgA-α1 anti-trypsin complex in the sera of patients with IgA myelomatosis. *Immunol. Lett.*, **11**, 277-280.

Sternberg, M. J. E. (1996). Protein structure prediction. IRL Press, Oxford.

Stewart, W. W. & Kerr, M. A. (1990). The specificity of the human IgA receptor (FcαR) determined by measurement of chemiluminescence induced by serum or secretory IgA1 or IgA2. *Immunol.*, **71**, 328-334.

Stockert, R. J. (1995). The asialoglycoprotein receptor: relationships between structure, function and expression. *Physiol. Rev.*, 75, 591-609.

Sumiyama, K., Saitou, N. & Ueda, S. (2002). Adaptive evolution of the IgA hinge region in primates. *Mol. Biol. Evol.*, **19**, 1093-1099.

Sun., Z., Almogren, A., Furtado, P. B., Chowdhury, B., Kerr, M. A. & Perkins, S. J. (2005). Semi-extended solution structure of human myeloma immunoglobulin D determined by constrained X-ray scattering. *J. Mol. Biol.*, **353**, 155-173.

Svehag, S-E. & Bloth, B. (1970). Ultrastructure of secretory and high-polymer serum immunoglobulin A of human and rabbit origin. *Science*, **168**, 847-849.

Takahashi, T., Iwase, T., Tachibana, T., Komiyama, K., Kobayashi, K., Chen, C. L., Mestecky, J. & Moro, I. (2000). Cloning and expression of the chicken immunoglobulin joining (J)-chain cDNA. *Immunogenetics*, **51**, 85-91.

Tomana, M., Niedermeier, W., Mestecky, J. & Skvaril, F. (1976). The differences in composition between the subclasses of IgA immunoglobulins. *Immunochem.*, **13**, 325-328.

Torano, A. & Putnam, F. W. (1978). Complete amino acid sequence of the α2 heavy chain of a human IgA2 immunoglobulin of the A2m(2) allotype. *Proc. Nat. Acad. Sci. USA.*, **75**, 966-969.

Underdown, B. J. & Schiff, J. M. (1986). Immunoglobulin A: strategic defence initiative at the mucosal surface. *Annu. Rev.*, **4**, 389-417.

Vaerman, J. P. 1971. Comparison between several mammalian IgAs, including the bovine. *J. Dairy Sci.*, **54**, 1317-8.

Vaerman, J. P, Langendries A, Vander MC. 1995. Homogenous IgA monomers, dimers, trimers and tetramers from the same IgA myeloma serum. *Immunol. Invest.*, **24**, 631-641.

van de Boog, P. J. M., van Zandbergen, G., de Fijter, J. W., Klar-Mohamad, N., van Seggelen, A., Brandtzaeg, P., Daha, M. R. & van Kooten, C. (2002). FcαRI/CD89 circulates in human serum covalently linked to IgA in a polymeric state. *J. Immunol.*, **168**, 1252-1258.

van Loghem, E. & Biewenga, J. (1983). Allotypic and isotypic aspects of human immunoglobulin A. *Mol. Immunol.*, **20**, 1001-1007.

van Munster, P. J. J. (1972). Secretory component. Thesis, University of Nijmegen.

Vogt, T. M., Blackwell, A. D., Giannetti, A. M., Bjorkman, P. J. & Enns, C. A. (2003) Heterotypic interactions between transferrin receptor and transferrin receptor 2. *Blood*, **101**, 2008-2014.

Volkov, V. V., Lapuk, V. A., Kayushina, R. L., Shytykova, E. V., Varlamova, E. Y., Malfois, M. & Svergun, D. I. (2003). Low-resolution structure of immunoglobulins IgG, IgM and rheumatoid factor IgM-RF from solution X-ray scattering data. *J. Appl. Cryst.*, **36**, 503-508.

Vladutiu, A. O. (2000). Immunoglobulin D: protperties, measurement and clinical relevance. *Clin. Diag. Lab. Immunol.*, **7**, 131-140.

Weltzin, R., Lucia-Jandris, P., Michetti, P., Fields, B. N., Kraehenbuhl, J. P. & Neutra, M. R. (1989). Binding and transepithelial transport of immunoglobulins by intestinal M cells: demonstration using monoclonal

IgA antibodies against enteric viral proteins. *J. Cell. Biol.*, **108**, 1673-1685.

Weiner, H. L. (1994). Oral tolerance. *Proc. Natl. Acad. Sci. USA*, **91**, 10762-10765.

White, K. D. & Capra J. D. (2002). Targeting mucosal sites by polymeric immunoglobulin receptor-directed peptides. *J. Exp. Med.*, **196**, 551-555.

Whitmore, A. C., Haughtton, G. & Arnold, L. W. (1992). *Annu. New York Acad. Sci.*, **651**, 143-151.

Wignall, G. D. & Bates, F. S. (1987). Absolute calibration of small-angle neutron scattering data. *J. Appl. Cryst.*, **20**, 28-40.

Williams, A. F. & Barclay, A. N. (1988). The immunoglobulin superfamily-domains for cell surface recognition. *Ann. Rev. Immunol.*, **6**, 381-405.

Wines, B. D., Hulett, M. D., Jamieson, G. P., Trist, H. M., Spratt, J. M. & Hogarth, P. M. (1999). Identification of residues in the first domain of Fcα receptor essential for interaction with IgA. *J. Immunol.*, **162**, 2416-2453 .

Woof, J. M., van Egmond, M. & Kerr, M. A. (2005). Fc receptors. In Mucosal Immunology 3rd Edition. Mestecky, J. R., Bienenstock, J., Lamm, M. E., Mayer, L., McGhee, J. R. & Strober, W. (eds). Academic Press, San Diego, 251–266.

Woof, J. M. & Mestecky, J. (2005). Mucosal immunoglobulins. *Immunological Rev.*, **206**, 64-82.

Woof, J & Kerr, M. A. (2006). The function of immunoglobulin A in immunity. *J. Pathol.*, **208**, 270-282.

Wu, T. T & Kabat. E. A. (1970). An analysis of the sequences of the variable regions of Bence-Jones proteins and myeloma light chains and their implications for body complementarity. *J. Exp. Med.*, **132**, 211-250.

Zikan, J., Novotny, J., Trapane, T. L., Koshland, M. E., Urry, D. W., Bennett, J. C. & Mestecky, J. (1985). Secondary structure of the immunoglobulin J chain. *Proc. Natl. Sci. USA*, **82**, 5905-5909.

Publications

Bonner, A., Perrier, C., Corthésy, B & Perkins, S. J. (2007). Solution structure of human secretory component and implication for biological function. *J. Biol. Chem.*, **282**, 16969-16980.

Bonner, A., Furtado, P. B., Almogren, A., Kerr, M. A. & Perkins, S. J. (2008). Implications of the near-planar solution structure of human myeloma dimeric IgA1 for mucosal immunity and IgA nephropathy. *J. Immunol.*, **180**, 1008-1018.

Perkins, S. J., Okemefuna, A. I., Fernando, A. N., Bonner, A., Gilbert, H. E. & Furtado, P. B. (2008). X-ray and neutron scattering data and their constrained molecular modelling. *Methods in cell Biol.*, **84**, 375-423.

Perkins, S. J. & Bonner, A. (2008). Structure determination of human and chimeric antibodies by solution scattering and constrained molecular modelling. *Biochem. Soc. Trans.*, **36**, 37-42.

Bonner, A., Almogren, A., Furtado, P. B., Kerr, M. A. & Perkins, S. J. (2008). Location of secretory component on the Fc edge of dimeric IgA1 reveals insight into the role of secretory IgA1 in mucosal immunity. *In Preparation.*

Bonner, A., Almogren, A., Furtado, P. B., Kerr, M. A. & Perkins, S. J. (2008). Implications for mucosal immunity of the non-planar IgA2 and near-planar IgA1 solution structures in human secretory IgA. *In preparation.*

Presentations and abstracts

Bonner, A., Gor, J., Perrier, C., Corthésy, B., & Perkins, S. J. Solution Structure of Recombinant Human Secretory Component by X-ray and Neutron Scattering and Analytical Ultracentrifugation. 14th International Symposium on Analytical Ultracentrifugation, Lausanne, Switzerland. 3rd – 4th March 2005 (Poster).

Bonner, A., Perrier, C., Corthésy, B., & Perkins, S. J. Solution Structure of Recombinant Human Secretory Component and its Two- and Three-Domain Fragments by Scattering, Ultracentrifugation and Constrained Modelling. 15th International Analytical Ultracentrifugation Symposium, University College London, London, U.K. 4th-7th April 2006 (Oral).

Bonner, A., Perrier, C., Corthésy, B., & Perkins, S. J. Solution Structure of Recombinant Human Secretory Component and its Two- and Three-Domain Fragments by Scattering, Ultracentrifugation and Constrained Modelling. 21st International Complement Workshop, Beijing, China. 22nd-28th October 2006 (Poster).

Bonner, A., Furtado, P. B., Almogren, A., Kerr, M. A. & Perkins, S. J. Planar solution structure of human dimeric IgA1 by X-ray and neutron scattering and implications for its function in humoral immunity. Neutrons in Biology, Rutherford Appleton Laboratory, Oxfordshire, U.K. 11th-13th July 2007 (Oral).

Perrier, C., Bonner, A., Perkins, S. J. & Corthésy, B. The three-dimensional structure of secretory component reveals the implication of surface glycans crucial for its microbial scavenger properties. International Congress of Mucosal Immunology Tokyo, Japan. 30th July 2007.

Perkins, S. J. & Bonner, A. Solution structure determinations of human and chimaeric antibodies by solution scattering, ultracentrifugation and constrained molecular modeling. Biochemical Society focused meeting, bringing together biomolecular simulation and experimental studies. Manchester Interdisciplinary Biocentre, U.K. 10th-11th September 2007.

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 282, NO. 23, pp. 16969–16980, June 8, 2007

Solution Structure of Human Secretory Component and Implications for Biological Function*[S]

Received for publication, February 12, 2007, and in revised form, March 9, 2007 Published, JBC Papers in Press, April 11, 2007, DOI 10.1074/jbc.M701281200

Alexandra Bonner[‡1], **Clémentine Perrier**[§2], **Blaise Corthésy**[§3], **and Stephen J. Perkins**[‡4]

From the [‡]*Department of Biochemistry and Molecular Biology, University College London, Darwin Building, Gower Street, London WC1E 6BT, United Kingdom and the* [§]*Service d'Immunologie et d'Allergie, Centre Hospitalier Universitaire Vaudois, BH19-650, Rue du Bugnon, CH-1011 Lausanne, Switzerland*

Secretory component (SC) in association with polymeric IgA (pIgA) forms secretory IgA, the major antibody active at mucosal surfaces. SC also exists in the free form, with innate-like neutralizing properties against pathogens. Free SC consists of five glycosylated variable (V)-type Ig domains (D1–D5), whose structure was determined by x-ray and neutron scattering, ultracentrifugation, and modeling. With a radius of gyration of 3.53–3.63 nm, a length of 12.5 nm, and a sedimentation coefficient of 4.0 S, SC possesses an unexpected compact structure. Constrained scattering modeling based on up to 13,000 trial models shows that SC adopts a J-shaped structure in which D4 and D5 are folded back against D2 and D3. The seven glycosylation sites are located on one side of SC, leaving known IgA-binding motifs free to interact with pIgA. This work represents the first analysis of the three-dimensional structure of full-length free SC and paves the way to a better understanding of the association between SC and its potential ligands, *i.e.* pIgA and pathogenic-associated motifs.

The mucosal epithelia lining the gastrointestinal, respiratory, and urogenital tracts represent the largest surface (400 m^2 for the human body) in contact with the external environment (1). Consequently, they are permanently exposed to foreign antigens and must therefore be efficiently protected to prevent possible damage by pathogenic agents (2). In response to this threat, a major immune defense system is mediated by secretory immunoglobulin A (SIgA)[5] within the lumen (3), where SIgA exerts its function of antigen neutralization. IgA is produced locally as polymeric IgA (pIgA), which consists of two or four IgA molecules linked covalently by a J chain (4). Subsequent transport of pIgA (and to a lesser extent pentameric IgM) across the epithelium is ensured by the polymeric Ig receptor (pIgR), which is expressed on the basolateral surface of epithelial cells. Following cleavage at luminal surfaces, SIgA is released as a complex of pIgA and the cleaved extracellular portion of the pIgR which is termed secretory component (SC). Because secretory IgM can be present in external secretions with an accessory role as a neutralizing antibody, this suggests that SC has a generic antibody-binding ability, being able to bind to both pIgA and pentameric IgM (5). This selectivity may be required to allow only larger, polymerized antibodies protected by heavily glycosylated SCs to enter the harsh environment of external secretions.

SC consists of five immunoglobulin-like domains (D1–D5) and up to seven glycan chains. SC bound to pIgA protects the antibody against proteolytic digestion (6) and governs anchoring of SIgA at mucosal surfaces (7). Free SC is also found in secretions and is now recognized as an active antibacterial participant to protect mucosal surfaces against *Helicobacter pylori* (8), enteropathogenic *Escherichia coli* and *Clostridium difficile* toxin A (9), and *Streptococcus pneumoniae* choline binding protein A (10, 11). Recombinant human SC produced from transfected Chinese hamster ovary (CHO) cells behaves identically, and its neutralizing properties rely on the presence of oligosaccharides on SC (12).

Despite the importance of SC and its glycosylation, only partial information is available on its structure. SC comprises the first 585 residues in pIgR (Fig. 1*A*) (13). Homology modeling of rabbit SC domains 1 and 2 (D1 and D2) indicates that both exhibit the typical features of Ig variable (V)-type superfamily members with seven β-strands A–G and an extra two designated C′ and C″ (14, 15). A similar topology was determined in the modeling of domains 2 and 3 (D2 and D3) of mouse SC (16). These were confirmed by the crystal structure of D1 of human SC, which revealed a V-type fold (17). So far, the three-dimensional domain arrangement of full-length SC has not been considered, even though this is essential to appreciate molecular mechanisms that are involved in mucosal immunity. Its high glycosylation and the presence of long interdomain linkers suggest that SC may be too flexible to be crystallized intact. In a situation such as this, x-ray and neutron scattering and analytical ultracentrifugation in combination with constrained modeling techniques will reveal the domain arrangement of SC (18). Here, we determine a solution structure for recombinant human SC produced in CHO cells. We show that SC forms a

* The authors have no conflicting financial interests. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 2OCW) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. S1–S4.

[1] Supported by the Biotechnology and Biological Sciences Research Council for financial support.
[2] Supported by a grant from the Nestlé Ph.D. Programme.
[3] Supported by the Swiss Science Research Foundation Grant 3200-109545.

[4] To whom correspondence should be addressed. Tel.: 44-20-7679-7048; Fax: 44-20-7679-7193; E-mail: s.perkins@medsch.ucl.ac.uk.

[5] The abbreviations used are: SIgA, secretory immunoglobulin A; SC, secretory component; Ig, immunoglobulin; pIgR, polymeric immunoglobulin receptor; CHO, Chinese hamster ovary; PBS, phosphate-buffered saline; V, variable type domain.

compact domain arrangement, and this structure rationalizes many of the biological roles of both free and bound SC for the first time.

MATERIALS AND METHODS

Recombinant Secretory Component and Its Fragments—CHO cells producing human SC (7) were cultured in suspension in CHO-S-SFM II medium (Invitrogen) using CELLine 350 bioreactors (Integra-Biosciences). Supernatants (8 ml) were collected twice a week, and immediately chromatographed onto a 1.6-cm × 1-meter long Superdex-200 column in PBS. Fractions containing human SC as confirmed by immunoblotting were pooled and filtered through 0.22-μm cartridges. Filtered aliquots were stored at 4 °C in PBS (116 mM NaCl, 10.4 mM Na_2HPO_4, 3.2 mM NaH_2PO_4, pH 7.2) prior to use for scattering and ultracentrifugation. Cleaved SC fragments were obtained from purified SC samples that were stored for at least 6 months at 4 °C. Separation was performed by chromatography onto a 1.6-cm × 1-m Superdex-200 column (Amersham Biosciences) in PBS. Pools of fractions containing fragments D1–D3 and D4–D5 were concentrated using Amicon Ultra (30-kDa molecular mass cut-off) and Centricon YM-10 (10-kDa molecular mass cut-off) filters (Millipore), respectively. Further purification of fragments D1–D3 and D4–D5 was achieved by repeating chromatography onto the Superdex-200 column. To identify the SC cleavage site, N-terminal sequencing of purified D4–D5 from a polyvinylidene difluoride blot was performed by Edman degradation at Alta Biosciences (University of Birmingham, UK).

For the x-ray scattering and ultracentrifugation experiments, samples were recovered in PBS. For neutron scattering, the buffer was Dulbecco's PBS (137 mM NaCl, 2.7 mM KCl, 8.1 mM Na_2HPO_4, 1.5 mM KH_2PO_4, pH 7.5), made up in 100% ^2H_2O. The neutron samples were dialyzed at 6 °C into this 100% ^2H_2O buffer for 36 h with four buffer changes. The two buffer densities in H_2O solvent were predicted to be 1.00485 g/ml and 1.00543 g/ml, respectively, from these compositions using SEDNTERP. The SC, D1–D3, and D4–D5 concentrations were determined using absorption coefficients of 12.3, 7.1, and 18.7 (1%, 280 nm, and 1-cm path length) that were calculated from its composition by assuming the presence of seven biantennary complex-type oligosaccharides $Man_3GlcNAc_4Gal_2NeuNAc_2$ (see Fig. 1) (19). The partial specific volume $\bar{\nu}$ for SC, D1–D3, and D4–D5 was calculated to be 0.717, 0.717, and 0.718 ml/g, respectively, from their compositions (19). The calculated molecular masses for SC, D1–D3, and D4–D5 were 79.6, 45.5, and 34.1 kDa, respectively, their unhydrated volumes were 99.7, 56.9, and 42.7 nm^3, respectively, and their hydrated volumes were 132.2, 75.5, and 56.7 nm^3, respectively (assuming 0.3 g of H_2O/g of glycoprotein).

X-ray and Neutron Scattering and Ultracentrifugation Data—X-ray scattering data were obtained in two sessions on the Beamline ID02 at the European Synchrotron Radiation Facility (Grenoble, France), operating at 6.03 GeV (20). To reduce the incident flux, the experiments were performed in 16-bunch and 4-bunch mode with storage ring currents ranging between 92 and 53 mA (SC) and 4-bunch mode with currents ranging between 41 and 23 mA (D1–D3 and D4–D5). The sample-to-detector distance of 3.0 m, yielded a Q range from 0.07 to 2.1 nm^{-1} (where $Q = 4\pi \sin\theta/\lambda$; 2θ = scattering angle; λ = wavelength). Samples were measured at concentrations between 0.62 and to 2.0 mg/ml (SC), 0.39 to 1.6 mg/ml (D1–D3), and 0.40 to 0.73 mg/ml (D4–D5). Data were acquired using five or ten time frames of 0.5, 1, 2, or 4 s to establish the absence of radiation damage. Neutron scattering data for SC were collected in one session in a Q range between 0.02 and 1.93 nm^{-1} on Instrument LOQ at the ISIS facility at the Rutherford Appleton laboratory, Didcot, UK (21). The proton beam current was 172–173 μA, and acquisition times were 1–2 h for concentrations between 0.62 and 2.0 mg/ml, measured at 15 °C. Other details, including calibrations and data reduction, the Guinier analyses to determine the radius of gyration R_G and the radius of gyration of the cross-sectional structure R_{XS}, and the calculation of the distance distribution function $P(r)$ using GNOM for which Q ranged between 0.08 and 2.06 nm^{-1}, are described previously (22, 23).

Analytical ultracentrifugation was performed at 20 °C on a Beckman XL-I instrument, equipped with an AnTi50 rotor. Sedimentation equilibrium data sets for SC between 0.20 and 2.50 mg/ml were acquired over 45 h with column heights of 2 mm at rotor speeds of 8,000, 11,000, 14,000, 18,000, and 23,000 rpm, until equilibrium had been reached at each speed as shown by the perfect overlay of runs measured at 5-h intervals. Sedimentation velocity data for SC, D1–D3, and D4–D5 were acquired over 16 h at rotor speeds of 20,000, 25,000, 30,000, and 35,000 rpm (and 42,000 rpm for D1–D3 and D4–D5) with column heights of 12 mm. SC was studied at five concentrations from 0.30 to 2.50 mg/ml. D1–D3 and D4–D5 were each studied at two concentrations of 0.39–1.6 mg/ml and 0.40–0.73 mg/ml respectively. DCDT+ $g(s^*)$ time-derivative analyses and SEDFIT Lamm equation fits based on the non-interacting species model were performed as previously described (23).

Constrained Modeling of SC—SC corresponds to residues 19–603 of the polymeric immunoglobulin receptor (SwissProt code P01833). The start and end of D2–D5 is defined as residues 1–104 in Fig. 1. The homology modeling of D2–D5 employed MODELLER v7 (24) and was based on a topological sequence alignment with the D1 crystal structure identified using BLASTP (PDB code 1xed) (17). The models were structurally validated using PROCHECK. Seven biantennary oligosaccharide chains $Man_3GlcNAc_4Gal_2NeuNAc_2$ were incorporated in extended conformations in the D1–D5 models (25, 23). The first SC residue Lys was added to the D1 crystal structure using MODELLER v7. To model the other interdomain linkers, residues 109–116, 219–221, 324–335, 440–446, and 543–585 (Fig. 1*B*) were created in extended β-strand conformations using INSIGHT II 98.0 with BIOPOLYMER, DISCOVER, HOMOLOGY, and DELPHI modules (Accelrys, San Diego, CA), and 5000 randomized conformations for each were generated using DISCOVER3 (22). Linker length constraints of 95–100% were applied to generate sufficient conformational variability when needed. Full SC models were created by superimpositions of the domains and the linkers.

Each SC, D1–D3, and D4–D5 model was converted into Debye spheres to calculate the x-ray and neutron scattering curves (22, 23). A cube side length of 0.542 nm in combination

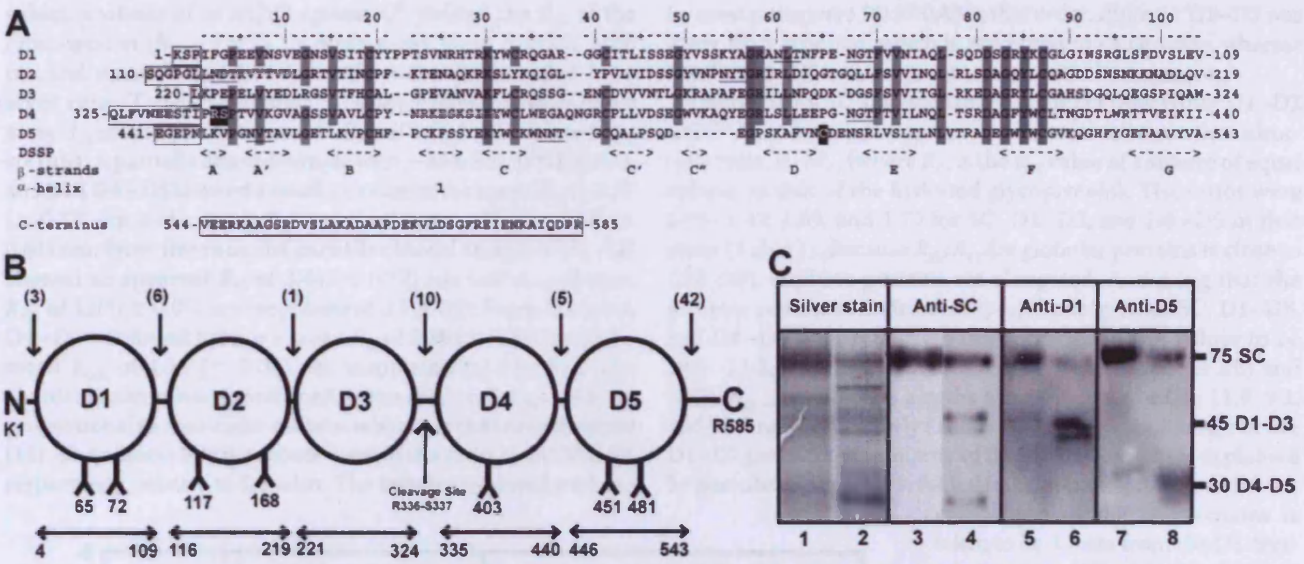

FIGURE 1. **The D1–D5 domains of human SC and characterization of recombinant polypeptides.** *A*, the sequence alignment of D2–D5 with D1. Conserved residues are highlighted in *gray*. The linker peptides are *boxed*. Putative glycosylation sites are *underlined*. The Arg[336]-Ser[337] cleavage site and Cys[502] that is covalently linked within SIgA are *highlighted* in *black*. The Define Secondary Structure Prediction-assigned secondary structure and the labeling of the β-strands are shown below the alignment. *B*, schematic representation of the V-type D1–D5 domains. Residue numberings indicate the cleavage site that yields D1–D3 and D4–D5 and the seven glycosylation sites. The lengths of linker peptides are shown in *brackets*. *C*, SDS-PAGE and Western blot analyses of SC (*lanes 1*, *3*, *5*, and *7*) and cleaved SC (*lanes 2*, *4*, *6*, and *8*). Silver stain (*lanes 1* and *2*) and Western blot performed with anti-SC (*lanes 3* and *4*), anti-D1 (*lanes 5* and *6*), and anti-D5 (*lanes 7* and *8*) identify full-length SC at 75 kDa, D1–D3 as a major band at 45 kDa, and D4–D5 at 30 kDa.

with a cutoff of 4 atoms consistently produced sphere models within 95% of the total unhydrated volume of 99.7, 56.9, and 43.8 nm^3, respectively, calculated from their compositions. Hence the optimal totals of unhydrated spheres were 626, 268, and 353 respectively, which were then hydrated. The number of spheres N in the unhydrated and hydrated models after grid transformation was used to assess steric overlap between the D1, D2, D3, D4, and D5 domains.

The models were assessed by calculation of the R_G and R_{XS} values in the same Q ranges used for the experimental Guinier fits. Models that passed these R_G and R_{XS} filters were ranked using the R-factor goodness-of-fit agreement. A flat background correction of 2.7% of $I(0)$ was applied to the final neutron scattering curve fits in Fig. 8 to allow for a uniform incoherent scattering of residual protons in the sample. Other details, including those of calibration studies used to validate this approach, are given elsewhere (22, 26). The sedimentation coefficients $s^0_{20,w}$ were calculated directly from the hydrated SC, D1–D3, and D4–D5 sphere models using the GENDIA and HYDRO programs (22, 23). Control calculations of $s^0_{20,w}$ were performed using HYDROPRO shell modeling directly from the molecular models for SC, D1–D3, and D4–D5 (27). The ten best SC α-carbon co-ordinate models were deposited in the Protein Data Bank with the accession code 2OCW.

RESULTS

Characterization of SC and Its Fragments—In SC, sequence similarities between D1 and the other four SC domains (D2–D5) show that all belong to the immunoglobulin (Ig) variable (V)-type classification (Fig. 1*A*). For clarification, the domains and interdomain linkers are depicted in Fig. 1*B*. As glycosylation and proper disulfide bonding are essential features of SC, we chose to produce the protein in CHO cells, which yields SC that is indistinguishable when compared with colostrum SC (12, 28). Purified recombinant human SC resulted in a major band at an apparent molecular mass of 75 kDa when analyzed by silver staining and immunodetection (Fig. 1*C*, *lanes 1*, *3*, *5*, and *7*). This molecular mass agrees well with the sequence-derived value of 79.6 kDa. Upon storage for several months at 4 °C, two additional bands at 45 and 30 kDa appear (Fig. 1*C*, *lanes 2*, *4*, *6*, and *8*). In Western blots, detection with polyclonal antisera specific for D1 and D5 established that the bands at 45 and 30 kDa represent cleavage of full-length SC into D1–D3 and D4–D5 fragments. This was confirmed by N-terminal sequencing of fragment D4–D5 that identified the sequence SPTVVKGVAG corresponding to the beginning of domain 4 and indicating cleavage between Arg[336] and Ser[337]. The PeptideCutter web tool (available from the ExPASy server) suggested that the enzymatic cleavage may be caused by one of trypsin, thrombin, or other related proteases. This unexpected observation provided us with three different polypeptides that were used to generate the results presented below.

X-ray and Neutron Scattering—X-ray scattering identified the domain arrangements in SC, D1–D3, and D4–D5 at concentrations between 0.40 and 2.7 mg/ml (see "Materials and Methods"). Because no radiation damage was detectable, the ten time frames of each acquisition were averaged for analyses. SDS-PAGE analyses before and after runs showed that, although SC and D4–D5 generally remained intact, further cleavages occurred in D1–D3 to give single domain fragments as well as the three-domain fragment. All the SC and D4–D5 results below correspond to samples shown to be intact by SDS-PAGE before and after the experiment.

At low Q values, Guinier analyses of ln $I(Q)$ against Q^2 resulted in linear plots for SC and its two fragments, yielding the radius of gyration (R_G) (Fig. 2, *A*, *E*, and *F*). At larger Q

values, analyses of ln $I(Q)Q$ against Q^2 yielded the R_G of the cross-section (R_{XS}). For SC, a mean x-ray R_G of 3.53 (± 0.03) nm and mean R_{XS} of 1.76 (± 0.08) nm were recorded from seven runs (Table 1). Another SC sample gave a similar mean x-ray R_G of 3.64 (± 0.17) nm and R_{XS} of 1.63 (± 0.12) nm from six runs. A partially cleaved sample with ~33% SC, 33% D1–D3, and 33% D4–D5 showed a small increase in the mean R_G to 3.77 (± 0.17) nm and a small decrease in the mean R_{XS} to 1.60 (± 0.61) nm. From five runs, the partially cleaved sample of D1–D3 showed an apparent R_G of 3.44 (± 0.27) nm and an apparent R_{XS} of 1.07 (± 0.07) nm (supplemental Fig. S1). From five runs, D4–D5 was found to have a mean R_G of 3.16 (± 0.04) nm and a mean R_{XS} of 1.11 (± 0.08) nm (supplemental Fig. S1). The Guinier analyses were confirmed by the $I(0)/c$ values, which are proportional to molecular masses, where c is the concentration (18). The observed $I(0)/c$ values were in the ratio of 1:0.30:0.53, respectively, relative to Lupolen. The sequence-derived molecular mass ratios were 1:0.57:0.43 in that order. $I(0)/c$ for D1–D3 was lower than expected, which is attributed to its cleavage, whereas that for D4–D5 was in fair agreement with the sequence.

Interestingly SC appeared more compact than either D1–D3 or D4–D5. Shape information was determined from the anisotropy ratio, R_G/R_O (where R_O is the R_G value of a sphere of equal volume to that of the hydrated glycoprotein). The ratios were 1.44–1.49, 1.69, and 1.70 for SC, D1–D3, and D4–D5 in that order (Table 1). Because R_G/R_O for globular proteins is close to 1.28 (29), all three proteins are elongated. Assuming that the proteins possess cylindrical shapes, the lengths of SC, D1–D3, and D4–D5 were calculated from its R_G and R_{XS} values to be 10.6–11.3, 11.3, and 10.2 nm, respectively. From the $I(0)$ and $[I(Q)Q]_{Q\to 0}$ values, the lengths were calculated to be 11.9, 9.1, and 10.8 nm, respectively (Table 1). Despite the cleavage in the D1–D3 sample, the similarity of these three lengths is explained by postulating that SC is folded back upon itself. If the end-to-end length of the SC domains is taken to be 3.7 nm from the D1 crystal structure (17), SC, D1–D3, and D4–D5 are predicted to be 18.5, 11.1, and 7.4 nm long if they are fully extended but with no linkers. The disparity between the observed SC length of 10.6–11.9 nm and the 18.5 nm prediction shows that SC must be folded back. That D4–D5 (Table 1) is observed to be longer than the predicted value may result from the 42 C-terminal residues in SC (Fig. 1*A*). The observed lengths for D1–D3 are in fair agreement with the predicted length of 11.1 nm.

Neutron scattering data for SC confirm the x-ray R_G values (Table 1). These were obtained as a control of radiation damage by x-rays, because these effects are absent in neutron scattering. The neutron data are complementary in that the hydration shell surrounding SC is not detectable (26). The high negative protein-solvent scattering contrast acts as a control of any scattering inhomogeneity effects caused by the high proportion of carbohydrate

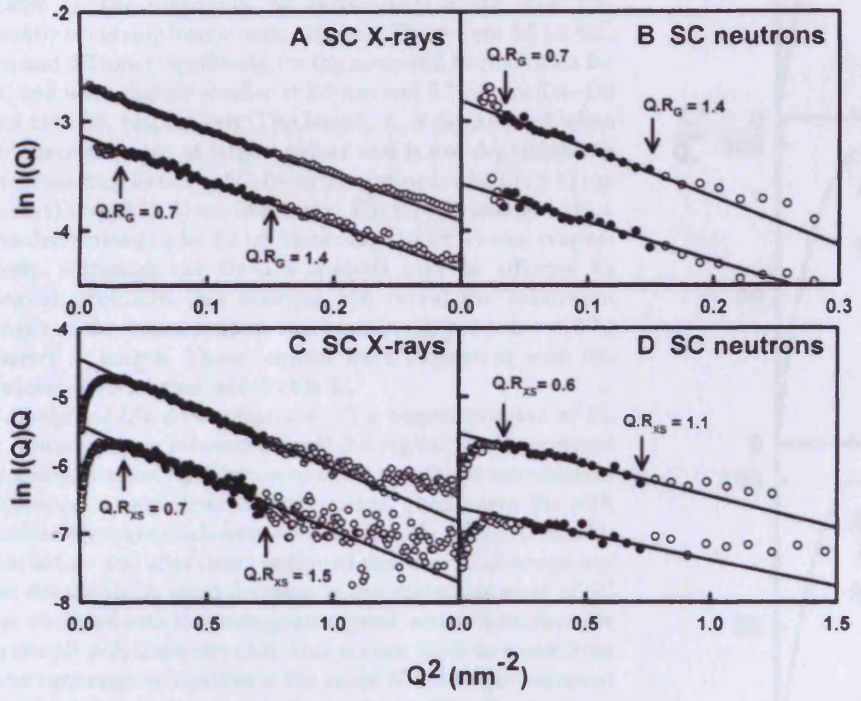

FIGURE 2. **Guinier R_G and R_{XS} analyses for SC.** Data points included in the Guinier fits are indicated by *filled circles* between the *arrowed* QR_G and QR_{XS} ranges. Where required for clarity, the Guinier plots were displaced on their vertical axes. The Q ranges for SC are 0.18–0.39 nm^{-1} (R_G) and 0.42–0.86 nm^{-1} (R_{XS}). *A* and *C*, x-ray fits for SC (1.3 mg/ml); *B* and *D*, neutron fits for SC (2.0 mg/ml (*upper*) and 1.3 mg/ml (*lower*)). For the Guinier analyses for D1–D3 and D4–D5, see supplemental Fig. S1.

TABLE 1
Experimental x-ray and neutron scattering and ultracentrifugation analyses for SC, D1–D3, and D4–D5

Protein	Guinier analyses					$P(r)$ analyses			AUC analyses		
	R_G	R_{XS}	R_G/R_0	L from R_G and R_{XS}	L from $I(0)$ and $[I(Q)Q]_{Q\to 0}$	R_G	M	L	$s^0_{20,w}$ (SEDFIT)	$s^0_{20,w}$ (DCDT+)	f/f_o
	nm					*nm*			*S*		
SC (x-ray)	3.53 ± 0.03	1.76 ± 0.08	1.44	10.6	11.9	3.66 ± 0.19	3.8 ± 0.2	13.0	4.2	4.0 (int)	1.55
SC (x-ray)	3.64 ± 0.17	1.63 ± 0.12	1.49	11.3	NA[a]	3.71 ± 0.12	NA	NA		3.9 (abs)	
SC (neutron)	3.63 ± 0.28	1.30 ± 0.10	NA	11.7	13.1	3.73 ± 0.17	3.2	12.0			
D1–D3 (x-ray)	3.44 ± 0.27[b]	1.07 ± 0.07[b]	1.69	11.3	9.1	3.52 ± 0.27	2.9	12.0	3.1	NA	1.42
D4–D5 (x-ray)	3.16 ± 0.04	1.11 ± 0.08	1.70	10.2	10.8	3.15 ± 0.18	2.7	10.0	2.4	2.4 (int) 2.3 (abs)	1.48

[a] NA, data not available.
[b] Apparent (see text).

Implications of the Near-Planar Solution Structure of Human Myeloma Dimeric IgA1 for Mucosal Immunity and IgA Nephropathy[1]

Alexandra Bonner,* Patricia B. Furtado,* Adel Almogren,† Michael A. Kerr,‡ and Stephen J. Perkins[2]*

IgA is unique in being able to form a diverse range of polymeric structures. Increases in the levels of dimeric IgA1 (dIgA1) in serum have been implicated in diseases such as IgA nephropathy. We have determined the solution structure for dIgA1 by synchrotron x-ray and neutron scattering and analytical ultracentrifugation. The Guinier radius of gyration (R_G) of 7.60–8.65 nm indicated that the two monomers within dIgA1 are arranged in an extended conformation. The distance distribution curve *P(r)* gave an overall length (*L*) of 22–26 nm. These results were confirmed by the sedimentation coefficient and frictional ratio of dIgA1. Constrained scattering modeling starting from the IgA1 monomer solution structure revealed a near-planar dimer structure for dIgA1. The two Fc regions form a slightly bent arrangement in which they form end-to-end contacts, and the J chain was located at this interface. This structure was refined by optimizing the position of the four Fab regions. From this, the best-fit solution structures show that the four Fab Ag-binding sites are independent of one another, and the two Fc regions are accessible to receptor binding. This arrangement allows dIgA1 to initiate specific immune responses by binding to FcαRI receptors, while still retaining Ag-binding ability, and to be selectively transported to mucosal surfaces by binding to the polymeric Ig receptor to form secretory IgA. A mechanism for the involvement of dIgA1 oligomers in the pathology of IgA nephropathy is discussed in the light of this near-planar structure. *The Journal of Immunology,* 2008, 180: 1008–1018.

Immunoglobulin A is the most abundant Ab class found in mucosal surfaces and is also the most heterogeneous human Ab class, being present in two subclasses IgA1 and IgA2 of which there are at least two different allotypes. This is reflected in the diverse functional properties of IgA, which are still not fully understood (1). Unlike any other Ig class, IgA occurs in different oligomeric states, circulating in serum predominantly as monomeric IgA at 3 mg/ml, and at mucosal surfaces as polymeric IgA, mainly as dimeric IgA covalently linked with secretory component (SC)[3] (2). In serum, IgA is the second most abundant Ab class, with ~90% being IgA1. Monomeric and dimeric IgA (dIgA) interact with the cell surface FcαR (FcαRI; CD89) to perform immune functions including the linking of the humoral and cellular immune response (3). In mucosal linings, lymphoid cells produce dIgA locally, which is secreted into the mucosa as secretory IgA (SIgA). Transport of dIgA across the epithelium into luminal secretions, and the consequent formation of SIgA is facilitated by the polymeric Ig receptor (pIgR). dIgA binds to pIgR at the basolateral surface of the epithelial cell, whereupon it is transcytosed across the cell and released into the lumen following cleavage of pIgR at the apical membrane. SIgA is released, with the extracellular pIgR domains (now called SC) remaining bound to the IgA dimer. The solution structure of SC has recently been elucidated (4).

An increased IgA level in serum can be associated with IgA nephropathy (IgAN) which is the most common form of chronic glomerulonephritis worldwide. IgAN is a renal disease involving the inflammation of the glomeruli in the kidneys, and is characterized by the mesangial deposition of polymeric IgA1. IgAN is caused by either an abnormality in IgA1 itself or in the production of IgA1, rather than an intrinsic abnormality in the glomerulus (5). Some reported characteristics of IgA in IgAN include abnormal *O*-glycosylation and an increased λ/κ L chain ratio (6). Serum from IgA myeloma patients can be rich in dIgA1 and polymeric IgA, although this does not lead to IgAN. In IgA myeloma and in other diseases characterized by high levels of dIgA1 and polymeric IgA, these forms of IgA are thought to play an important role in some of the pathology associated with these diseases (7).

Polymeric IgA consists of two (or sometimes three or four) IgA monomers bound covalently through a joining chain (J chain) (Fig. 1). Monomeric IgA has a 12-domain structure arranged as two H and two L chains (Fig. 1). Crystal structures are known for the murine Fab region of IgA and the human Fc region of IgA1 bound to FcαRI (8, 9). The topology of each of the IgA domains is a β-sheet sandwich comprised of the DEBA and GFC β-strands for the seven-stranded C-type domains and the DEBA and GFCC′C″ β-strands for the nine-stranded V-type domain (10). IgA1 and IgA2 differ significantly in the hinge region where IgA2 lacks a Pro-, Ser-, and Thr-rich sequence of 13 aa that is present in IgA1

*Department of Biochemistry and Molecular Biology, University College London, London, United Kingdom; †Department of Pathology, College of Medicine and King Khalid University Hospital, King Saud University, Riyadh, Saudi Arabia; and ‡Department of Clinical Biochemistry and Immunology, General Infirmary, Leeds, United Kingdom

Received for publication September 17, 2007. Accepted for publication November 9, 2007.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was supported by the Biotechnology and Biological Sciences Research Council.

[2] Address correspondence and reprint requests to Dr. Stephen J. Perkins, Department of Biochemistry and Molecular Biology, Darwin Building, University College London, Gower Street, London WC1E 6BT, U.K. E-mail address: s.perkins@medsch.ucl.ac.uk

[3] Abbreviations used in this paper: SC, secretory component; dIgA1, dimeric IgA1; SIgA, secretory IgA; pIgR, polymeric Ig receptor; IgAN, IgA nephropathy; HSA, human serum albumin.

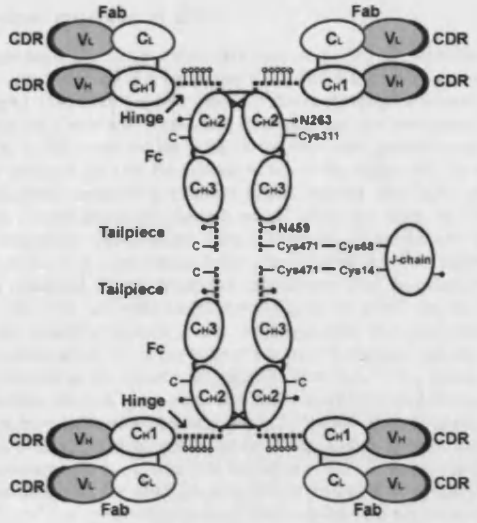

FIGURE 1. The domain structure and sequence of dIgA1. dIgA1 is formed by two IgA1 monomers that are covalently linked with the J chain. Each IgA1 H chain contains the V_H, C_H1, C_H2, and C_H3 domains, and each L chain contains the V_L and C_L domains. The V and C domains are highlighted with gray and white backgrounds, respectively. The CDRs (black crescents), the 23-residue hinge peptide (dashed line), and the 18-residue C-terminal tailpiece (dashed line) are highlighted. Inter-H chain disulphide bridges at Cys^{241}-Cys^{241} and two at Cys^{242}-Cys^{299} are shown as three black lines. Cys^{471} in one tailpiece of each Fc fragment is disulphide-bridged with either Cys^{14} or Cys^{68} in the J chain. The *N*-linked oligosaccharide sites at Asn^{263} and Asn^{459} are denoted by filled symbols (●). The approximately five *O*-linked oligosaccharides in each hinge are denoted by open symbols (○).

(11, 12). IgA2 also differs in that in the commonest allotype (IgA2m(1)), the L chains are not covalently linked to the H chains but to each other, and IgA2 is held together only by noncovalent interactions (13). Both IgA subclasses have an additional 18-aa C-terminal peptide on each H chain called the tailpiece (Fig. 1). Cys^{471} in the tailpiece of the two IgA monomers forms a disulphide bridge with the J chain (14). Thus, the tailpiece is important for the dimerization of IgA1 and IgA2 (15). The J chain is an 18-kDa Ig-related glycoprotein containing 137 residues and one *N*-linked carbohydrate.

The three-dimensional structures of the different forms of polymeric IgA are not well-understood. Structures for monomeric human IgA1 and IgA2 together with the covalent adduct of IgA1 with human serum albumin (HSA) have been determined by constrained scattering modeling (16–18). The overall structure of human myeloma dIgA has been observed by electron microscopy as a tail-to-tail arrangement of the Fc regions in the two IgA monomers (19–22). However, the conditions of electron microscopy work using heavy metal stains in vacuo can be harsh, and the solution arrangement of the monomers in dIgA1 was unknown up to now. The size of dIgA1, together with its high glycosylation and long flexible hinge and tailpiece regions, suggests that its crystal structure will be hard to achieve. In such a situation, x-ray and neutron scattering and analytical ultracentrifugation combined with constrained modeling, will result in a medium resolution structure determination that will reveal molecular details within dIgA (23). By this, we show that myeloma dIgA1 is formed as a near-planar molecule in solution with a bent arrangement of Fc regions. This structure rationalizes the accommodation of the J chain within dIgA1, as well as effector functions of dIgA1, particularly its interaction with FcαRI receptors and SC in relation to the formation of SIgA. Insight is also provided into the role of dIgA1 in the development of IgAN.

Materials and Methods

Preparation and composition of dIgA1

A monoclonal dIgA1 with κ-class L chain was isolated from a human myeloma serum using a combination of thiophilic chromatography and jacalin-agarose lectin affinity chromatography (24, 25). Samples were subjected to size-exclusion chromatography to remove nonspecific aggregates and checked by reducing and nonreducing SDS-PAGE to confirm sample integrities before and after data collection. For x-ray scattering and analytical ultracentrifugation experiments, dIgA1 was dialyzed into Dulbecco's PBS supplemented with EDTA and sodium azide as antibacterial preservatives (12.5 mM sodium phosphate, 140 mM NaCl, 0.5 mM EDTA, 0.02% NaN_3, pH 7.4). For neutron scattering, the buffer was Dulbecco's PBS as above, dialyzed at 6°C into 100% ^2H_2O for 36 h with four buffer changes. The dIgA1 amino acid and carbohydrate composition for data analyses corresponds to two IgA1 molecules as previously described (16). The human J chain sequence was taken from the SWISSPROT sequence code P01591, to which another biantennary complex-type oligosaccharide was added at Asn^{49}. This resulted in the following: molecular mass, 344.8 kDa, an unhydrated volume of 439.4 nm^3, a hydrated volume of 580.2 nm^3, an absorption coefficient at 280 nm of 12.3 (1%, 1 cm), and a partial specific volume $\bar{v}$ of 0.724 ml/g.

X-ray and neutron scattering

Solution scattering is a diffraction technique that studies the overall structure of biological macromolecules in random orientations in solution (23). X-ray scattering data were obtained on the Beamline ID02 at the European Synchrotron Radiation Facility (ESRF; Grenoble, France) with a ring energy of 6.0 GeV in single-bunch mode with currents from 11 to 12 mA to reduce the incident x-ray flux. The sample-to-detector distance of 3.0 m yielded a Q range from 0.07 nm^{-1} to 2.1 nm^{-1} ($Q = 4\ \pi \sin\ \theta/\lambda$; 2θ = scattering angle; λ = wavelength). Samples were contained in water-cooled Perspex cells at 15°C, of path thickness 1 mm and mica windows of thickness 25 μm at concentrations of 0.58 and 1.16 mg/ml. Samples were measured in 10 time frames, each of 2 s, these being optimal to eliminate radiation damage effects. Neutron scattering data were obtained on Instrument LOQ in two beam sessions at the pulsed neutron source ISIS at the Rutherford Appleton Laboratory. Neutrons were derived from proton beam currents of ~180 μA. Samples and buffers were measured in 2-mm-thick rectangular quartz Hellma cells positioned in a thermostated rack at 15°C. Data acquisitions lasted 4–13 h using concentrations of 0.83 and 1.16 mg/ml. Other details, including calibrations and data reduction, the Guinier analyses to determine the radius of gyration R_G and the radius of gyration of the cross-sectional structure R_{XS}, and the calculation of the distance distribution function $P(r)$ using GNOM, are described previously (17). For GNOM, the dIgA1 x-ray $I(Q)$ curve contained 524 data points between Q values of 0.98 and 2.02 nm^{-1} and was fitted with a D_{max} set as 26 nm. The dIgA1 neutron $I(Q)$ curve contained 61 data points between Q values of 0.21–2.1 nm^{-1} and was fitted with a D_{max} set as 23.5 nm.

Analytical ultracentrifugation

Analytical ultracentrifugation studies macromolecular structures in solution by following their sedimentation behavior on subjecting these to a high centrifugal force (23). Sedimentation equilibrium data were acquired over 45 h using six-sector cells with column heights of 2 mm at rotor speeds of 5,000, 8,000, 11,000, 14,000, 17,000, and 20,000 rpm at 20°C on a Beckman XL-I instrument equipped with an AnTi50 rotor. Absorbance scans at 280 nm and interference scans were recorded at six concentrations between 0.07 and 0.97 mg/ml. Molecular weights were determined on the assumption of a single species using ORIGIN version 4.1 (Microcal). The buffer density was calculated to be 1.00543 g/ml and the viscosity was taken as 0.01002 cp. Sedimentation velocity data were acquired at concentrations of 0.53 and 1.00 mg/ml over 16 h at rotor speeds of 10,000, 15,000, and 20,000 rpm in two-sector cells with column heights of 12 mm. Data were analyzed using DCDT$^+$ $g(s^*)$ time-derivative analyses and SEDFIT v9.4 based on the continuous $c(s)$ distribution model with a resolution set as 150, while the cell meniscus and bottom, the frictional ratio of 1.638, and the partial specific volume $\bar{v}$, buffer density and viscosity were held fixed, and the baseline was allowed to float. Other details are as described previously (17).

Constrained modeling of dIgA1

The scattering and sedimentation data were subjected to constrained modeling to determine solution structures (23). The full coordinates of monomeric IgA1 (16) was used to generate dimer models. A planar starting model for the dimer was created by arranging the two monomers back to back. The J chain was not included. A dummy atom position was defined midway between the two Fc regions to act as the origin for the rotations and translations required to generate dIgA1 models. One IgA1 monomer was then rotated about the dummy atom, while the other monomer remained stationary. The rotations were in steps of 10° between 0° and 180° on each of the *x*-, *y*-, and *z*-axes, hence this resulted in 19^3 or 6859 models for each rotational search. Monomer separations were assessed by setting this as 1.40, 2.78, and 4.60 nm in three searches in which the two starting monomers shared a common *y*-axis. This separation was that between the two Cα atoms of Lys^{457} at the base of the two Fc regions, and the dummy atom is located at the mean position of all four Lys^{454} Cα atoms. In two final searches, the two IgA1 monomers had a separation of 2.78 nm, but the IgA1 monomer to be rotated was translated 2.0 and 4.0 nm along the *x*-axis.

The coordinate models were converted into Debye spheres to compute their scattering curves. A cube side length of 0.531 nm and a cutoff of four atoms consistently gave models within 2% of their total unhydrated volume of 417.2 nm^3, i.e., that without the J chain as this had not been modeled, and gave a total of 2787 unhydrated spheres. After hydration, the number of hydrated spheres *n* was 3678. The values of *n* and the modeled R_G, $R_{XS\text{-}1}$, and $R_{XS\text{-}2}$ parameters were calculated in the same *Q* ranges used for the experimental Guinier analyses. Models that fit within 5% of *n* and the x-ray Guinier values were ranked using the goodness-of-fit *R* factor of the model. A flat-background correction of 0.5% of *I(0)* was applied to the neutron fits to allow for residual incoherent scattering. The sedimentation coefficients $s°_{20,w}$ were calculated from the coordinate models using HYDROPRO version 7c. To take into account the hydration shell surrounding dIgA1, the recommended value of 0.31 nm for the atomic element radius for all atoms was used as an empirical approximation of this. Other details, including those of calibration studies that validate these modeling methods, are given elsewhere (16, 17). The 10 best-fit dIgA1 α-carbon coordinate models were deposited in the Protein Data Bank with the accession code 2QTJ. The α-carbon coordinate model of the related pentameric human IgM structure was deposited with the accession code 2RCJ (24).

Results

Purification of serum dIgA1

Conventional purifications of polymeric IgA from normal human serum are inadequate for purifying the polymeric forms of IgA. A procedure based on thiophilic resins with a binding capacity of over 28 mg of Ig/ml of resin allows rapid purification of Igs which bind selectively to the resin while almost all the other serum proteins do not (25, 26). dIgA1 was purified from anonymized myeloma serum selected for its high polymer content. The myeloma serum had an Ig content of IgA at 15 g/L with a dimer to monomer ratio of 2:1, while IgG was present at 7g/L and IgM was at 1.5 g/L. dIgA1 was shown by nonreducing SDS-PAGE to be well-resolved from monomeric and trimeric IgA1 species in subsequent gel filtration (26). Its characterization is described in detail elsewhere (26). Reducing SDS-PAGE showed the expected bands corresponding to the IgA1 H chain at 62 kDa, and the L chain at 30 kDa. Monoclonal anti-κ L chain and anti-human J chain Abs were used in Western blots to confirm the presence of dIgA1. In Sephacryl S300 size-exclusion gel filtration, dIgA1 eluted as a single symmetrical peak, which was concentrated for data collection.

X-ray and neutron scattering

The averaged solution arrangement of the monomers within dIgA1 was analyzed by comparing x-ray scattering data *I(Q)* for dIgA1 with previous x-ray data for serum IgA (16, 17). Data were collected at concentrations between 0.6 and 1.2 mg/ml at the ESRF. The comparison of data from a single time-frame exposure with the average over 10 consecutive time frames showed that the Guinier analyses were unaffected by the exposure time, indicating that radiation damage or x-ray induced aggregation were absent. The time-averaged runs were thus used for data analyses. At the lowest *Q* values, these analyses resulted in linear plots in *Q* ranges of 0.10–0.17 nm^{-1} that gave R_G values within satisfactory $Q.R_G$ limits (Fig. 2). The x-ray R_G value for dIgA1 was 8.65 ± 0.27 nm (eight values) (Fig. 2*A*). This is significantly larger than the R_G value of 6.20 ± 0.13 nm for monomeric IgA1 (16) and 5.18 ± 0.09 nm for monomeric IgA2m(1) (17). The anisotropy ratio R_G:R_O (where R_O is the R_G value of the sphere with the same volume as the hydrated glycoprotein) for dIgA1 was 2.16, where that for monomeric IgA1 was 1.99, that for monomeric IgA2 was 1.66 and that for the IgA1-HSA complex was 2.13 (16–18). This showed that dimer formation had increased the degree of elongation of the IgA1 structure to form a more extended one.

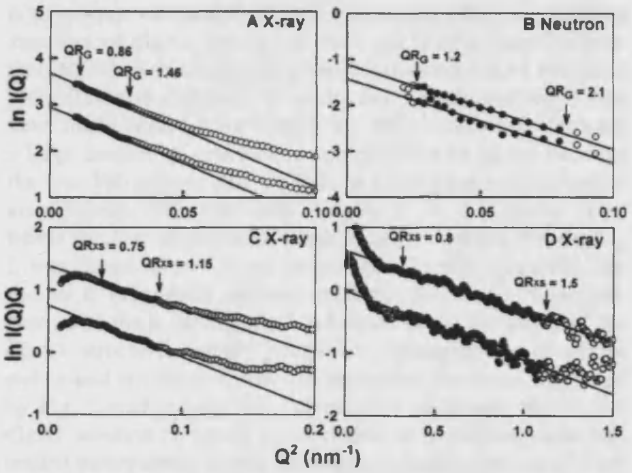

FIGURE 2. X-ray and neutron Guinier analyses for dIgA1. The $Q.R_G$ and $Q.R_{XS}$ ranges used to determine the R_G, $R_{XS\text{-}1}$, and $R_{XS\text{-}2}$ values are represented by filled circles between the arrowed data points. *A* and *B*, The x-ray and neutron *Q* ranges used for the R_G analyses were 0.10–0.17 nm^{-1}. *C* and *D*, The *Q* ranges used to calculate the $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ values were 0.20–0.28 nm^{-1} and 0.56–1.04 nm^{-1}, respectively. In *A*, *C*, and *D*, the x-ray fits for dIgA1 correspond to concentrations of 1.16 (*upper*) and 0.58 mg/ml (*lower*). In *B*, the neutron fits for dIgA1 correspond to 1.16 (*upper*) and 0.83 mg/ml (*lower*).

The shorter axes of dIgA1 were monitored by x-ray cross-sectional Guinier analyses that resulted in the $R_{XS\text{-}1}$ and $R_{XS\text{-}2}$ parameters. The *ln (I(Q).Q)* vs Q^2 plots showed two linear regions in *Q* ranges of 0.20–0.28 nm^{-1} and 0.56 to 1.04 nm^{-1} (Fig. 2, *C* and *D*). The $R_{XS\text{-}1}$ value of 3.94 ± 0.18 nm for dIgA1 is almost double that for monomeric IgA1, IgA2, and the IgA1-HSA complex which are 2.20, 2.47, and 2.31 nm, respectively (16–18). This indicates that the IgA1 dimer is significantly different in shape from the monomer. The $R_{XS\text{-}2}$ value of 1.43 ± 0.07 nm is close to the $R_{XS\text{-}2}$ of 1.56, 1.47, and 1.31 nm for IgA1, IgA2, and IgA1-HSA (16–18). The similarity of these four $R_{XS\text{-}2}$ values suggests that the Fab regions retain their distinct identities within the monomeric and dimeric structures. If dIgA1 is crudely approximated as an elliptical cylinder in shape, the combination of the R_G and $R_{XS\text{-}1}$ values resulted in an overall length *L* for dIgA1 of 25 nm. The comparison of this length with that of 20.1 nm for IgA1 and 16.1 nm for IgA2 (16, 17) showed that the increased size of dIgA1 does not correspond to a doubling in length of the monomers.

Complementary neutron scattering data was obtained for dIgA1 at 0.6–1.5 mg/ml in 100% ^2H_2O buffer as a control for x-ray-induced radiation damage. This also acts as a control of internal scattering inhomogeneity effects caused by the estimated 9.6% carbohydrate content of dIgA1. A high negative protein-solvent scattering contrast is now used in place of the high positive contrast observed by x-rays. Neutron scattering also provided a different

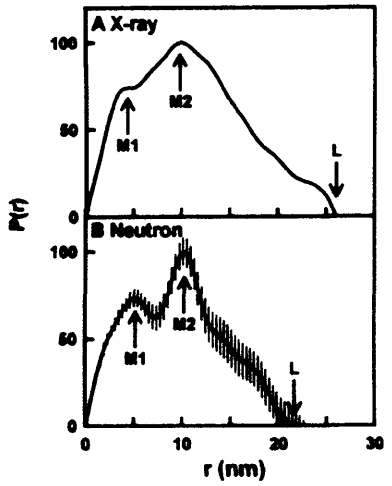

FIGURE 3. X-ray and neutron distance distribution functions *P(r)* for dIgA1. *A* and *B* show the x-ray and neutron *P(r)* curves respectively. The concentrations of dIgA1 were 1.15 (x-rays) and 1.16 mg/ml (neutrons). The maxima *M1* and *M2* of the *P(r)* curve depict two frequently occurring distances within the dIgA1 structure. The positions of *M1* and *M2* are 4.94 and 9.88 nm, respectively (x-rays), and 5.06 and 10.14 nm, respectively (neutrons). The length of IgA1 is denoted by *L* at the *r* value where the *P(r)* curve reaches 0. The *L* values for dIgA1 were 26 (x-rays) and 22 nm (neutrons).

view of dIgA1 in that the hydration shell surrounding dIgA1 is visible by x-ray scattering, but not by neutron scattering (27). The neutron R_G value of dIgA1 was 7.60 ± 0.05 nm (two values) (Fig. 2*B*). This value supports the x-ray R_G determination. However, this is 1 nm less than the x-ray R_G value, which is a greater difference than that seen previously with IgA1 and IgA2 (16, 17). This is attributed to the positioning of most of the hydration shell at large distances from the centre of the dIgA1 structure. The neutron anisotropy ratio R_G:R_O was 2.08, and confirmed the x-ray value. As the consequence of poor counting statistics at large *Q*, no neutron $R_{XS\text{-}1}$ or $R_{XS\text{-}2}$ values were determined for dIgA1. Neutrons also lead to the molecular mass, as there is a linear relationship between the LOQ Guinier *I(0)/c* values for proteins measured in ^2H_2O buffers and molecular mass, where molecular mass = $I(0)/c \times 9.10^5$ (16). The dIgA1 neutron Guinier *I(0)/c* value of 0.33 ± 0.05 resulted in a molecular mass value of 297 ± 45 kDa. This measurement is within error of the composition-derived value of 345 kDa for dIgA1, showing that the scattering data is consistent with the expected composition of dIgA1 with 9 *N*-linked and 20 *O*-linked oligosaccharides (Fig. 1).

Structural dimensions are provided by the transformation of the *I(Q)* curve into the distance distribution function *P(r)* curve. The mean R_G values determined from the x-ray and neutron *P(r)* curves were 8.67 ± 0.17 nm (six values) and 7.47 nm (two values), respectively. These were consistent with the Guinier R_G values. The x-ray *P(r)* curve for dIgA1 was of better quality than the neutron *P(r)* curve for reason of better signal-noise ratios (Fig. 3), nonetheless the two curves were reproducible in that a double peak was observed with maxima of *M1* and *M2* that were located at similar r values of 4.9–5.1 and 9.9–10.1 nm, respectively. These two peaks correspond to an abundance of interatomic vectors within the dIgA1 solution structure. The peak *M1* at 4.9–5.1 nm was also evident in the *P(r)* curve for monomeric IgA1 and IgA2 and for myeloma IgA1-HSA at 3.7, 4.5, and 4.9 nm, respectively (16–18). *M1* is assigned to the most commonly occurring distance within a single Fab or Fc region, each of approximate length 8 nm. Hence, *M1* in Fig. 3 is consistent with a well-defined Fab region within the solution structure of dIgA1. *M2* at 9.9–10.1 nm is of greater intensity than *M1* when this is compared with monomeric IgA1 and IgA2 (16, 17). In the absence of molecular modeling (below), the most likely reason for this intensity difference is that there are a large number of interatomic vectors close to 10 nm between the four Fab regions that are held in a comparatively inflexible arrangement. The maximum length, *L*, is determined from where the *P(r)* curve reaches zero at large *r* values. For dIgA1, *L* was found to be 26 nm (x-ray) and 23 nm (neutron). The shorter L value with neutrons indicates that a significant proportion of the hydration shell is located at the periphery of the dIgA1 structure. Simple geometric considerations of planar end-to-end arrangements of two monomer structures suggested by Fig. 1 readily lead to *L* values close to 26 nm. Hence, the dIgA1 solution structure is concluded to be formed as an extended arrangement of two monomers joined at the end of their Fc regions, and not as a more compact solution structure that might resemble a bouquet of four Fab flowers positioned on two Fc stalks.

Analytical ultracentrifugation

To confirm the molecular mass of dIgA1, sedimentation equilibrium experiments were performed at six concentrations between 0.07 and 0.97 mg/ml and at six rotor speeds using both interference optics and absorbance optics at 280 nm (*Materials and Methods*). On the assumption that a single species was present, sedimentation equilibrium fits were conducted with nine curves at each dIgA1 concentration using both data types. The individual fits gave random fit residuals and molecular mass values that ranged between 280 and 345 kDa. Their extrapolation to zero concentration gave a mean molecular mass of 335 (±15) kDa (data not shown). These are within error of the molecular mass of 345 kDa calculated from the sequence of dIgA1 if the 29 glycosylation sites in this are all occupied by biantennary oligosaccharides. Had the glycosylation of dIgA1 corresponded to 29 tetra-antennary oligosaccharides, the molecular mass would have been 357 kDa. This outcome confirmed the neutron molecular mass determination (see above).

The sedimentation coefficient $s^\circ_{20,w}$ provides an independent measure of macromolecular elongation to the R_G value. For dIgA1, sedimentation velocity experiments were performed at three speeds at two concentrations (*Materials and Methods*). The *c(s)* distribution plots using SEDFIT software identify the species present. All the experimental and fitted dIgA1 boundaries showed good visual agreement with satisfactory root mean square deviation values (Fig. 4, *A* and *B*). The resulting *c(s)* distributions in Fig. 4, *C* and *D*, showed that only a single species of dIgA1 was present with a $s^\circ_{20,w}$ value of 9.7 S. The conversion of the *c(s)* plots to molecular mass distributions *c(M)* showed that this corresponded to a molecular mass of 342 (±11) kDa. This agreed well with the sedimentation equilibrium value of 335 (±15) kDa and sequence-derived molecular mass of 345 kDa. No higher oligomers were detected in the *c(s)* plots. An analysis based on 4–20 scans at the center of the sedimentation profile using the *g(s*)* method in DCDT$^+$ software confirmed the SEDFIT results. Here, the dIgA1 scans gave a $s^\circ_{20,w}$ value of 9.8 S (interference) and 9.9 S (absorbance) (Fig. 5). The SEDFIT and DCDT$^+$ determinations agree well with previous literature reporting a $s^\circ_{20,w}$ value of 10 S (28), 9.65 S (21), and 9.2–9.3 S (29). The $s^\circ_{20,w}$ value leads to the frictional ratio f/f_O, where f_O is the *f* value of the sphere with the same volume as the hydrated glycoprotein. Its value for dIgA1 of 1.63 is slightly larger than those of 1.56 found for monomeric IgA1 and 1.53 for monomeric IgA2 (17). This

Streptococcus species in the same way as normal dIgA1 (Ref. 26, A. Almogren and M. A. Kerr, unpublished data). On SDS-PAGE, the myeloma dIgA1 had the same mobility as normal dIgA1 (26). Functionally, the myeloma dIgA1 triggered neutrophil respiratory burst through CD89 in the same way as normal dIgA1, and bound to purified SC. Hence, our new structure of dIgA will aid our understanding and subsequent ability to manipulate the mucosal immune system and clarify the pathology of diseases. The size of dIgA1, its high carbohydrate content and its flexibility mean that the crystallization of intact dIgA1 may not be realistic, thus constrained scattering modeling presently offers the only route to a molecular structure. Experimentally, we have shown that dIgA1 has a near-planar structure in solution with the two monomers orientated end-to-end through their Fc regions. The two Fc regions are slightly bent in their relative arrangement (Fig. 9). Because the dIgA1 structure was determined by fitting models against experimental scattering data, this is not a prediction method, and the final models qualify for deposition in the Protein Data Bank.

Although scattering modeling is not able to determine unique molecular structures, it is able to rule out poorer-fit models (Table I). Both search cycles successfully converged on one small family of related best-fit structures. One striking outcome is the close end-to-end proximity of the two Fc regions in dIgA1. The J chain does not act as a spacer, as sometimes suggested in cartoons of the IgA dimer. Likewise, the two Fc regions do not overlap as discussed in earlier reviews (32). A bent Fc arrangement was modeled, which has also been noted by electron microscopy (19). The length of the Fc-Fc region within dIgA1 in Fig. 9*B* is close to 14.4 nm. This agrees well with the 14.0–15.5 nm, 13.8 (±2.6) nm and 12.5 nm determinations from electron microscopy (19–21, 28). The J chain is best accommodated in dIgA1 within a crevice formed by the bent Fc regions, although a crystal structure will be needed to confirm this. If the J chain had separated the two Fc regions, this would add 2.0–2.5 nm to the length of the Fc-Fc regions. This outcome was ruled out by our Searches 2 and 3. Our location of the J chain in dIgA1 is consistent with our scattering modeling of pentameric IgM (24). There, the J chain was best fitted within the planar disc of the five IgM Fc regions, and not as a spacer. The early electron microscopy studies of dIgA1 indicated flexible planar end-to-end structures (20–22, 28), but it was not known whether these near-planar structures exist in solution. The constrained modeling showed that the best-fit dIgA1 structures are near planar. Interestingly a near-planar dIgA1 structure would be consistent with the planar IgM structure (24). Models showing perpendicular arrangements of the two monomers or the four Fab regions were eliminated during the rejection of poor-fit models (Figs. 6, *B–D*, and 7, *B–D*).

The solution structures for monomeric IgA1, the IgA1-HSA complex and dIgA1 all demonstrated similar T-shaped structures for the two Fab regions relative to the Fc region (16, 18). A well-separated Fab-Fc structure in all three cases was deduced from the observation of the *M2* peak in the *Pr* curves, meaning that the two Fab regions are consistently seen on average to be extended away from the Fc region. Hinge flexibility of the Fab region relative to the Fc region is nonetheless expected, as observed from the variable conformations seen in electron microscopy studies of dIgA1, and the altered position of the Fab regions on going from the monomer to the dimer (Figs. 7*K* and 9*A*). This is in accordance with the ability of IgA1 to bind efficiently to a range of Ags. The maintenance of these extended hinge conformations can be rationalized in terms of its primary structure (PVPSTPPTPSPPTPPT PSPSCCH). Even though the hinge is comparatively long at 23 residues, the presence of *O*-linked NeuNAc.Gal.GalNAc trisaccharides, multiple prolines and the absence of glycines are expected to maintain their extended conformations. Interestingly, human IgD also possesses an *O*-glycosylated segment within its 64-residue hinge, and this has a similar T-shaped structure to that for IgA1 (24). As the four Fab regions are well separated within dIgA1, each Ag-binding site at the CDRs (CDR in Fig. 9*B*) is potentially able to interact with repeated epitopes independently of each other. The valency of dIgA1 is therefore predicted to be four, and its avidity is predicted to be higher compared with its monomer. This assumes that the linearity of the dIgA1 structure has no bearing on its capacity to bind independently to four Ag molecules.

A near-planar dIgA1 structure would be well-adapted for interactions with its receptors. By this, dIgA1 is able to approach the host cell membrane surface for effective interactions with a cell surface receptor, most notably FcαRI which is responsible for IgA-mediated phagocytosis, oxidative burst, and Ab-dependent cellular cytoxicity among other roles (33, 34). The dIgA1 model shows that receptor binding would occur independently of Ag binding. The dIgA1 residues involved in binding to FcαRI are highlighted in orange in Fig. 9*B* (9, 35). By superimposition of the Fc crystal structure complexed with two FcαRI molecules (9) upon the dIgA1 model, it is seen that the two FcαRI receptors will project out of the plane of dIgA1 but on opposite sides of this dIgA1 plane (green in Fig. 9, *C* and *D*). The superimposition of a second crystal structure onto the other monomer within dIgA1 shows likewise that the third and fourth FcαRI receptors project outward on both sides of this plane (yellow in Figs. 9, *C* and *D*). Previously it has been shown that two FcαRI receptors can bind simultaneously to the monomeric Fc region in solution, and it has been speculated that this bivalent binding occurs at cell surfaces (9, 17, 36). Fig. 9, *C* and *D*, likewise leads to the speculation that dIgA1 is able to bind up to two FcαRI receptors on a given cell surface, even though its predicted valency for receptors is four. The spatial separation between the C-terminal α-carbon atoms of the two FcαRI receptors bound to monomeric and dimeric IgA1 is similar at 12.3 and 10.2 nm, respectively, meaning that the same arrangement of FcαRI receptors will theoretically bind equally well to monomers or dimmers of IgA1. However, for reason of comparatively low receptor binding affinities in relation to the IgA1 serum concentrations, it seems more likely that a 1:1 stoichiometry results in vivo (33, 34).

The appearance of dIgA1 as a near-planar structure with a bent double Fc region leads to a possible mechanism for the formation of SIgA1 from dIgA1 and SC. The bend between the two Fc regions in the final models of Fig. 6 offers a covalently bound location for the J chain within the crevice that is created by this bend. This location may enable the J chain in dIgA1 to be sufficiently exposed to bind to a central region in SC, this being the first event in the selective transportation of dIgA1 across epithelial cells to form SIgA (37). The interaction between the C-terminal region of the J chain in dIgA1 and SC is noncovalent and represents a critical step for transport of this complex to mucosal surfaces to form SIgA (37). A "zipper effect" binding of SC to dIgA1 has been proposed, where the N-terminal D1 domain of SC first interacts noncovalently with dIgA1 (possibly with the Fc region), a central region in SC then interacts with the C-terminal region of the J chain, and dIgA1 and SC are brought into the correct position for covalent binding between Cys^{311} in the C_H2 domain of dIgA1 and Cys^{502} in the D5 domain of pIgR. A J-shaped solution structure for free SC has been recently determined (4). It is possible that the center of this may form an initial contact with the J chain in dIgA1, then SC will unfold toward both ends of dIgA1 as the D1 and D5 domains interact successively across the two Fc regions. The wrapping of the D1–D5 domains around the Fc regions in dIgA1 then

provides SIgA1 with a greater ability to resist proteolysis in the harsh environment of mucosal surfaces.

The near-planar dIgA1 structure provides insight into the formation of other polymeric IgAs. The structure of dimeric IgA2 is predicted to be similar to that of dIgA1 because the two monomers are joined through the base of their two Fc regions to the J chain and there are only minor sequence differences between the IgA1 and IgA2 isotypes (11). The hinge region is not involved in the formation of the IgA2 dimer. However, the binding site of SC to dimeric IgA2 differs from that to dIgA1, as SC binds covalently to dIgA1 but may not do so to dimeric IgA2 (18). The two dIgA isotypes may therefore differ in a manner that remains to be elucidated, most probably mediated by the closer proximity of the Fab regions to SC in secretory IgA2. The antigenic reach of the two isotypes will differ for reason of the different hinge structures, as previously discussed (16, 17), and dimer formation is expected to accentuate this difference. The formation of trimeric and tetrameric IgA structures (38) can be visualized by analogy with the structure of planar pentameric IgM in which its Fc regions are disulphide linked through the Cys^{414} residues in the Cμ3 domains and the Cys^{575} residues in the tailpieces (24). Hexameric IgM structures without J chain have been described in which a sixth IgM monomer is inserted within this plane (39). Cys^{311} in IgA1 (Fig. 1) is located at the identical sequence and structural position to that of Cys^{414} in IgM (9). By analogy with IgM, it is relatively straightforward to insert further monomers within this Fc plane to form larger IgA oligomers. Using molecular graphics based on the IgA1 Fc crystal structure and the IgM scattering structure, it was possible to create trimeric (and tetrameric) assemblies of three (and four) coplanar Fc regions starting from the dimer with the J chain present, but only if the Fab regions are absent. If the Fab regions and hinges are present, they will readily lead to significant steric conflicts with the assembly of Fc trimers and tetramers of IgA1 unless these Fab regions are significantly displaced out of the plane containing the three or four Fc regions.

IgAN is defined by the deposition of IgA in the glomerular mesangium of the kidney and this is predominantly polymeric IgA1 (5). A high level of polymeric IgA1 such as in IgA myelomas does not necessarily lead to IgAN. A reduced galactosylation of the hinge *O*-linked oligosaccharides has been implicated in the pathology of IgAN. It is possible that this change may destabilize the extended conformation of the IgA1 hinge and cause the Fab regions to be displaced. Hence glycosylation changes may provide one possible mechanism for IgA aggregate formation by permitting the Fc regions to self-associate. Another possibility for IgA aggregation arises from the consequences of the cleavage of the IgA1 hinge by bacterial proteases. This was seen to cause the cleaved Fc fragments derived from dimeric and trimer IgA1 (but not monomeric IgA1) to form disulphide-linked multimers (26). Our molecular graphics modeling of planar Fc rings of IgA by analogy with IgM (see above) lead to the hypothesis that, if the IgA1 hinges become more flexible through deglycosylation and cause the Fab regions to move away, or the Fab regions are partially or completely removed by bacterial proteolysis, then near-planar assemblies of trimers, tetramers, and higher oligomers of IgA can be generated starting from the double Fc region of dIgA1 and its J chain. If this hypothesis is correct, these Fc oligomers can be stabilized by disulphide bridges through Cys^{311} at the surface of the C_H2 domain in IgA, and may lead to the IgA deposits seen in IgAN.

The elucidation of a near-planar solution structure of dIgA1 has yielded a better understanding of the function of polymeric forms of IgA. This opens the way for the constrained modeling of SIgA1 and SIgA2. Knowledge of the monomer arrangement within the dimer has provided insights into the assembly of this and higher oligomers of IgA1, and the way that this potentially interacts with the FcαRI receptor and SC. However, many unsolved issues remain, including the exact binding motifs within dIgA1 that interact with SC in the formation of SIgA, and the structure and composition of pathological dIgA1 forms that are implicated in diseases such as IgAN.

Acknowledgments

We thank Dr. R. E. Saunders for excellent computational support, J. Gor for ultracentrifugation support, Dr. S. Finet (ESRF, Grenoble, France) and Drs. R. K. Heenan and S. M. King (ISIS, Rutherford Appleton Laboratory) for instrumental support, and Dr. A. Robertson for assistance with the initial scattering data collection.

Disclosures

The authors have no financial conflict of interest.

References

1. Russell, M. W., M. Kilian, and M. E. Lamm. 1999. Biological activities of IgA. In *Mucosal Immunology*, 2nd Ed. P. L. Ogra, J. Mestecky, E. Lamm, W. Strober, J. Bienenstock, and J. R. McGhee, eds. Academic Press, San Diego, pp. 225–240.
2. Kerr, M. A. 1990. The structure and function of human IgA. *Biochem. J.* 271: 285–296.
3. Morton, H. C., M. van Egmond, and J. G. J. van de Winkel. 1996. Structure and function of human IgA Fc receptors (FcαR). *Crit. Rev. Immunol.* 16: 423–440.
4. Bonner, A., C. Perrier, B. Corthésy, and S. J. Perkins. 2007. The solution structure of human secretory component and implications for biological function. *J. Biol. Chem.* 282: 16969–16980.
5. Barratt, J., and J. Feehally. 2005. IgA nephropathy. *J. Am. Soc. Nephrol.* 16: 2088–2097.
6. Coppo, R., and A. Amore. 2004. Aberrant glycosylation in IgA nephropathy. *Kidney Int.* 65: 1544–1547.
7. Mestecky, J., H. Suzuki, T. Yanagihara, Z. Moldoveanu, M. Tomana, K. Matousovic, B. A. Julian, and J. Novak. 2007. IgA nephropathy: current views of immune complex formation. *Contrib. Nephrol.* 157: 56–63.
8. Suh, S. W., T. N. Bhat, M. A. Navia, G. H. Cohen, D. N. Rao, S. Rudikoff, and D. R. Davies. 1986. The galactan-binding immunoglobulin Fab J539: an X-ray diffraction study at 2.6 Å resolution. *Proteins* 1: 74–80.
9. Herr, A. B., E. R. Ballister, and P. J. Bjorkman. 2003. Insights into IgA-mediated immune responses from the crystal structures of human FcαRI and its complex with IgA1-Fc. *Nature* 423: 614–620.
10. Chothia, C., and E. Y. Jones. 1997. The molecular structure of cell adhesion molecules. *Annu. Rev. Biochem.* 66: 823–862.
11. Torano, A., and F. W. Putnam. 1978. Complete amino acid sequence of the α 2 heavy chain of a human IgA2 immunoglobulin of the A2m (2) allotype. *Proc. Natl. Acad. Sci. USA* 75: 966–969.
12. Putnam, F. W., Y. S. V. Lui, and T. L. K. Low. 1979. Primary structure of a human IgA1 immunoglobulin. *J. Biol. Chem.* 254: 2865–2874.
13. Grey, H. M., C. A. Abel, W. J. Yount, and H. G. Kunkel. 1968. A subclass of human γA-globulins (γA2) which lacks the disulphide bonds linking heavy and light chains. *J. Exp. Med.* 128: 1223–1236.
14. Garcia-Pardo, A., M. E. Lamm, A. G. Plaut, and B. Frangione. 1981. J chain is covalently bound to both monomer subunits in human secretory IgA. *J. Biol. Chem.* 256: 11734–11738.
15. Atkin, J. D., R. J. Pleass, R. J. Owens, and J. M. Woof. 1996. Mutagenesis of the human IgA1 heavy chain tailpiece that prevents dimer assembly. *J. Immunol.* 157: 156–159.
16. Boehm, M. K., J. M. Woof, M. A. Kerr, and S. J. Perkins. 1999. The Fab and Fc fragments of IgA1 exhibit a different arrangement from that in IgG: a study by X-ray and neutron solution scattering and homology modelling. *J. Mol. Biol.* 286: 1421–1447.
17. Furtado, P. B., P. W. Whitty, A. Robertson, J. T. Eaton, A. Almogren, M. A. Kerr, J. M. Woof, and S. J. Perkins. 2004. Solution structure determination of human IgA2 by X-ray and neutron scattering and analytical ultracentrifugation and constrained modelling: a comparison with human IgA1. *J. Mol. Biol.* 338: 921–941.
18. Almogren, A., P. B. Furtado, Z. Sun, S. J. Perkins, and M. A. Kerr. 2006. Biochemical and structural properties of the complex between human immunoglobulin A1 and human serum albumin by X-ray and neutron scattering and analytical ultracentrifugation. *J. Mol. Biol.* 356: 413–431.
19. Bloth, B., and S.-E. Svehag. 1971. Further studies on the ultrastructure of dimeric IgA of human origin. *J. Exp. Med.* 133: 1035–1042.
20. Munn, E. A., A. Feinstein, and A. J. Munro. 1971. Electron microscope examination of free IgA molecules and of their complexes with antigen. *Nature* 231: 527–529.
21. Feinstein, A., E. A. Munn, and N. E. Richardson. 1971. The three-dimensional conformation of M and A globulin molecules. *Ann. NY Acad. Sci.* 190: 104–121.
22. Nezlin, R. 1998. *The Immunoglobulins: Structure and Function.* Academic Press, New York.
23. Perkins, S. J., A. I. Okemefuna, A. N. Fernando, A. Bonner, H. E. Gilbert, and P. B. Furtado. 2008. X-ray and neutron scattering data and their constrained molecular modelling. *Methods Cell Biol.* 84: 375–423.

24. Perkins, S. J., A. Nealis, B. J. Sutton, and A. Feinstein. 1991. Solution structure of human and mouse immunoglobulin M by synchrotron X-ray scattering and molecular graphics modelling: a possible mechanism for complement activation. *J. Mol. Biol.* 221: 1345–1366.
25. Sun, Z., A. Almogren, P. B. Furtado, B. Chowdhury, M. A. Kerr, and S. J. Perkins. 2005. Semi-extended solution structure of human myeloma immunoglobulin D determined by constrained X-ray scattering. *J. Mol. Biol.* 353: 155–173.
26. Almogren, A., and M. A. Kerr. 2008. Irreversible aggregation of the Fc fragment derived from polymeric but not monomeric serum IgA1- implications in IgA-mediated disease. *Mol. Immunol.* 45: 87–94.
27. Perkins, S. J. 2001. X-ray and neutron scattering analyses of hydration shells: a molecular interpretation based on sequence predictions and modelling fits. *Biophys. Chem.* 93: 129–139.
28. Dourmashkin, R. R., G. Virella, and R. M. E. Parkhouse. 1971. Electron microscopy of human and mouse myeloma serum IgA. *J. Mol. Biol.* 56: 207–208.
29. Bjork, I., and E. Lindh. 1974. Gross conformation of human secretory immunoglobulin A and its component parts. *Eur. J. Biochem.* 45: 135–145.
30. Zikan, J., J. Novotny, T. L. Trapane, M. E. Koshland, D. W. Urry, J. C. Bennett, and J. Mestecky. 1985. Secondary structure of the immunoglobulin J chain. *Proc. Natl. Acad. Sci. USA* 82: 5905–5909.
31. Frutiger, S., G. J. Hughes, N. Paquet, R. Lüthy, and J. C. Jaton. 1992. Disulphide bond assignment in human J chain and its covalent pairing with IgM. *Biochemistry* 31: 12643–12647.
32. Mestecky, J., and J. R. McGhee. 1987. Immunoglobulin A (IgA): molecular and cellular interactions involved in IgA biosynthesis and immune response. *Adv. Immunol.* 40: 153–183.
33. Monteiro, R. C., and J. G. van de Winkle. 2003. IgA Fc receptors. *Annu. Rev. Immunol.* 21: 177–204.
34. Woof, J. M., M. van Egmond, and M. A. Kerr. 2005. Fc receptors. In *Mucosal Immunology*, 3rd Ed. J. R. Mestecky, J. Bienenstock, M. E. Lamm, L. Mayer, J. R. McGhee, and W. Strober, eds. Academic Press, San Diego, pp. 251–266.
35. Pleass, R. J., J. I. Dunlop, C. M. Anderson, and J. M. Woof. 1999. Identification of residues in the CH2/CH3 domain interface of IgA essential for interaction with the human Fcα receptor (FcαRI) CD89. *J. Biol. Chem.* 274: 23508–23514.
36. Herr, A. B., C. L. White, C. Milburn, C. Wu, and P. J. Bjorkman. 2003. Bivalent binding of IgA1 to FcαRI suggests a mechanism for cytokine activation of IgA phagocytosis. *J. Mol. Biol.* 327: 645–657.
37. Johansen, F.-E., R. Braathen, and P. Brandzaeg. 2001. The J chain is essential for polymeric Ig receptor-mediated epithelial transport of IgA. *J. Immunol.* 167: 5185–5192.
38. Mestecky, J., I. Moro, M. A. Kerr, and J. M. Woof. 2005. Mucosal immunoglobulins. In *Mucosal Immunology*, 3rd Ed. J. R. Mestecky, J. Bienenstock, M. E. Lamm, L. Mayer, J. R. McGhee, J. R. and W. Strober, eds. Academic Press, San Diego, pp. 153–181.
39. Davis, A. C., K. H. Roux, and M. J. Shulman. 1988. On the structure of polymeric IgM. *Eur. J. Immunol.* 18: 1001–1008.